DNA Replication Origins in *Haloferax volcanii*

Michelle Sarah Hawkins, MBiol.

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy,
February 2009

Contents

Abstract

DNA replication is fundamental to the proliferation of life. Sites of DNA replication initiation are called replication origins. Bacteria replicate from a single origin whereas eukaryotes utilise multiple origins for each chromosome. The archaeal domain includes species which replicate using multiple origins of replication in addition to those which use a single origin. Archaeal DNA replication proteins are similar to eukaryotic replication machinery. Most characterised archaeal origins are adjacent to an *orc* gene which encodes a homologue of the Orc1 subunit of the eukaryotic initiator protein complex.

Replication origins of the halophilic archaeon *Haloferax volcanii* were identified using a combination of genetic, biochemical and bioinformatic approaches. *H. volcanii* has a multireplicon genome consisting of a circular main chromosome and three mini-chromosomes: pHV1, pHV3 and pHV4. The major chromosome contains multiple origins of replication and is the first example of multiple origins on a single replicon in the Euryarchaeota. Each characterised origin is adjacent to an *orc* gene and contains repeated sequence motifs surrounding an A/T-rich duplex unwinding element.

The archaeal recombinase, RadA, is homologous to eukaryotic and bacterial Rad51/RecA. It is widely held that deletion of *radA* results in elimination of homologous recombination. In this study the discovery of a *radA*-independent recombination pathway specific to replication origins is described. This dynamic mechanism was identified by observing chromosomal integration of plasmids containing *H. volcanii* replication origins in a *radA* deletion strain.

The eukaryotic *RAD25* gene is involved in nucleotide excision repair and transcription. *H. volcanii* has four *RAD25* homologues, one on pHV4 and three near the *oriC-2* locus on the main chromosome. A role for the assistance of *oriC-2* firing is proposed based on autonomously replicating plasmid assays. Deletion of all four *RAD25* homologues did not increase DNA damage sensitivity but resulted in a minor growth defect.

Acknowledgements

Firstly I would like to thank my supervisor Thorsten Allers for his guidance, training and ideas. Most of all I am indebted to him for giving me the chance to work on a fascinating project.

I would like to take this opportunity to thank everyone who has contributed to the results presented in this thesis, particularly Cédric Norais and Hannu Myllykallio.

Everyone in the laboratory has had an impact on my PhD and I want to thank all past and present members. I have thoroughly enjoyed working in C5 and I will miss all its idiosyncrasies. I would like to give special thanks to Sam, Stéphane, Belinda, Lee and Roxane for their advice and stimulating conversations, both in and out of science.

The University Trampoline club has been an immensely positive force during the last three years. After all, it is hard to worry about stubborn cloning while somersaulting! To Viki, Jo, Nick and every other bouncer-thanks for everything. Trampolining also introduced me to Bernie who has been a great friend. Thank you Bernie for all your advice, the good times and even the jogging! Thank you to Dan for being my best friend and sharing my love of gymnastics. I am very grateful to my family for supporting me in my studies and giving me the chance to get this far.

My deepest thanks go to Matt for his love, patience, and support throughout our time together. I am proud to now call him my husband and he deserves most of the credit for keeping me sane!

Finally I would like to thank *Haloferax volcanii* for not being straightforward and therefore making me a better scientist.

Abbreviations

2-D ..... two-dimensional
5-FOA ..... 5-fluoroorotic Acid
ACS ..... autonomously replicating sequence consensus sequence
ARS ..... autonomously replicating sequence
BER ..... base excision repair
BLAST ..... basic local alignment search tool
bp ..... base pair
CDK ..... cyclin-dependent kinases
Chr. ..... chromosome
CFU ..... colony forming unit
ChIP ..... chromatin immunoprecipitation
COG ..... cluster of orthologous groups
cSDR ..... constitutive stable DNA replication
DDK ..... Dbf4-dependent kinases
DHFR ..... dihydrofolate reductase
DNA ..... deoxyribose nucleic acid
dNTP ..... deoxynucleotide
DSBR ..... double-stranded break repair
dsDNA ..... double-stranded deoxyribose nucleic acid
DUE ..... duplex unwinding element
EDTA ..... ethylenediaminetetraacetic acid
HMG-CoA ..... 3-hydroxy-3-methylglutaryl coenzyme A
HTH ..... helix-turn-helix
Hv-Ca ..... *Haloferax volcanii* casamino acids media
Hv-YPC ..... *Haloferax volcanii* yeast extract, peptone and casamino acids media
iSDR ..... induced stable DNA replication
kb ..... kilobase
Mb ..... megabase
MCM ..... mini-chromosome maintenance
mev ..... mevinolin
MMC ..... mitomycin C
MMS ..... methanesulfonic acid methyl ester/methyl methanesulfonate

mORB ..... mini-origin recognition box
NER ..... nucleotide excision repair
NEF ..... nucleotide excision repair factor protein assemblies
OBR ..... origin of bidirectional replication
ORB ..... origin recognition box
ORC ..... origin recognition complex
ORF ..... open reading frame
PCNA ..... proliferating cell nuclear antigen
PCR ..... polymerase chain reaction
PFG ..... pulsed-field gel electrophoresis
RDR ..... recombination-dependent replication
RIP ..... replication initiation point mapping
rRNA ..... ribosomal deoxyribose nucleic acid
SDSA ..... synthesis-dependent strand annealing
SSB ..... single-strand binding protein
ssDNA ..... single-stranded deoxyribose nucleic acid
TIGR ..... The Institute for Genomic Research
TFIIH ..... transcription factor II H
trp ..... tryptophan
ura ..... uracil
UV ..... ultraviolet light
v/v ..... volume per volume
w/v ..... weight per volume
WH ..... winged helix
WhiP ..... winged-helix initiator protein
WT ..... wild-type
X-gal ..... X-galactosidase
XP ..... xeroderma pigmentosum

Chapter 1: Introduction

1.1 DNA replication

DNA replication is crucial to life. Without DNA replication genetic information can not be transmitted to the next generation. To replicate the entire genome before cell division the stages of initiation, elongation and termination of DNA replication must be accurately regulated.

Diverse mechanisms have evolved to ensure that DNA initiation events are spatially and temporally coordinated and lead to complete replication within a single cell cycle. To achieve this it might be advantageous to initiate DNA replication from specific locations rather than begin DNA replication at random points in the genome. In 1963 Jacob, Brenner and Cuzin proposed the "replicon model" to explain how this could be regulated (Jacob and Brenner, 1963). The replicon model proposed that *trans*-acting factors called initiators, bound specific *cis*-acting sequences called replicators. Binding of the initiator to the replicator promoted DNA replication initiation from the replicator sequence.

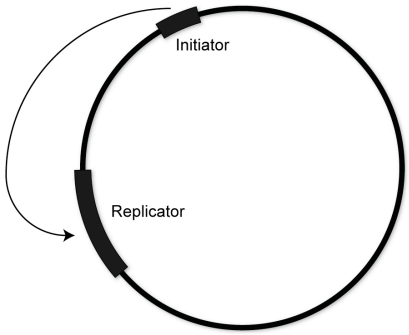


Figure 1.1: Replicon model
Replicon model applied to a circular bacterial chromosome.

The replicon model has been a useful paradigm for studying DNA replication in bacteria but it has been less successful when applied to eukaryotes. Replication start sites are called "DNA replication origins" and the chain of events from initiation at an origin to replication fork establishment can be summarised as follows:

1. Initiator binds and marks origin.
2. Initiator remodels DNA architecture and recruits replication factors e.g. replicative helicase.
3. Helicase activity unwinds the DNA duplex.
4. Local unwinding creates an open DNA structure allowing replisome formation. (Kornberg, 1992)

The replisome is a dynamic multi-protein complex that performs DNA replication at replication forks. When all the components of the replisome have been assembled, the bi-directional elongation phase can begin. When elongation is completed fork termination occurs and finally concatenated DNA molecules are separated to resolve daughter chromosomes. Our understanding of origins and initiation regulation varies considerably depending upon the model system. In the following description of DNA replication the focus will be on replicator characteristics and initiator proteins in a variety of organisms.

1.2 Bacterial DNA replication

Bacteria have a single origin called *oriC* and the vast majority of bacteria have circular chromosomes. Replication forks proceed bidirectionally from *oriC* until termination. The most studied bacterial origin is the *oriC* of *Escherichia coli* but around a dozen bacterial species have had their origins experimentally mapped. Sequence conservation of *oriC* is limited outside of closely related species but the features of origins are conserved across bacteria. This enables bioinformatic identification of origins in species with sequenced genomes and (Mackiewicz *et al.,* 2004). In general, *oriC* is a 200-1000 bp intergenic region that frequently lies within the gene cluster: *rnpA-rmpH-dnaA-dnaN-recF-gyrB-rnpA*. These genes encode proteins that function in DNA recombination and processing. *oriC* is usually adjacent to *dnaA*, the bacterial initiator (Kornberg, 1992). Concordance of initiator and replicator may be advantageous in enabling DnaA to bind *oriC* as soon as it is translated, it is also suggestive of co-regulation. Alternatively this synteny may have been preserved to prevent disruption of the *dnaA-oriC* system. The proximity of replicator and initiator in bacterial genomes is a clear example of the replicon model. DnaA binds to *oriC* at multiple asymmetric 9 bp DnaA boxes. The consensus *E. coli* DnaA box sequence is: 5′-TTATNCACA (Messer, 2002). The DnaA box is highly

conserved. Bacterial chromosomes contain multiple DnaA boxes, several of which cluster within *oriC*. Another sequence feature of *oriC* is an A/T-rich region. A-T nucleotides base pair by forming two hydrogen bonds while G-C base pairs bond with three. A-T-rich DNA is therefore inherently less stable and easier to unwind and melt. Origin melting via these A/T-rich duplex unwinding elements (DUE) is conserved amongst bacteria (Mackiewicz *et al.,* 2004). Binding of DnaA induces a sharp curvature in the DNA duplex (Schaper and Messer, 1995). This bending of the origin-DnaA complex facilitates unwinding of the DUE and subsequent helicase loading.

DnaA performs many roles during initiation. In addition to recognising the origin and promoting duplex unwinding, DnaA also guides the replicative helicase (DnaB) and the helicase loader (DnaC) complex to its loading site (Messer, 2002). DnaA can also act as a transcription factor by binding to DnaA boxes in promoters, for example DnaA can repress *uvrB* (DNA repair) and *rpoH* (RNA polymerase factor) gene expression (Messer and Weigel, 1997). Given these diverse functions the DnaA protein has been divided in to four functional domains (Messer *et al.,* 1999; Sutton and Kaguni, 1997):

I. N-terminal module for DnaB recruitment and DnaA oligomerisation.
II. Extended linker segment (highly variable between species).
III. AAA+ family ATP binding cassette for DUE unwinding.
IV. C-terminal DNA binding domain for DnaA box recognition.

Bacterial origins of replication have been studied in depth using *E. coli* as a model system. Therefore, most of the biochemical and genetic data used to formulate the current model of initiation is from *E. coli.* Structural information for DNA complexed with *E. coli* DnaA domain IV (Fujikawa *et al.,* 2003), and the crystal structure of *Aquifex aeolicus* DnaA (Erzberger *et al.,* 2002) supports the functional domain divisions. The first step of initiation is the recognition of *oriC* by DnaA, this occurs by binding of domain IV to the DnaA box. Both structures show that the basis of DnaA sequence- specific recognition is the insertion of part of the domain IV helix-turn-helix (HTH) motif in to the major groove of the DnaA box. One monomer of DnaA binds to each DnaA box and this induces a 40° bend in the DNA duplex (Schaper and Messer, 1995). Positions 2, 4, 7 and 9 of the consensus sequence are particularly important for DnaA binding and are all positioned on one side of the

DNA helix (Speck *et al.,* 1997). The 3′ 5 bases are bound by van der Waals contacts with methyl groups of thymine bases and the 5′ 3 bases of the recognition sequence are bound by base specific hydrogen bonds (Fujikawa *et al.,* 2003). This sequence specificity results in a high affinity interaction. A conserved basic loop proximal to the HTH motif is also important for DNA binding and may mediate minor groove interactions (Erzberger *et al.,* 2002). DnaA is an AAA+ ATPase and domain III alters its conformation depending on whether ADP or ATP is bound (Erzberger *et al.,* 2006). ATP/ADP-DnaA complexes bind DnaA boxes with the same affinity but DnaA complexed with ATP also binds *oriC* with a low affinity at 6-mer sites common in the DUE (consensus sequence 5′-AGATCT) (Speck *et al.,* 1999). This lower affinity binding is cooperative with the DnaA box binding (Speck and Messer, 2001) and only ATP-DnaA binding induces DUE remodelling. Since the DnaA boxes and ATP-DnaA recognition sites are clustered, DnaA binding localises multiple monomers to *oriC*, which in turn leads to oligomerisation. The DnaA crystal structure failed to indicate potential DnaA oligomer interfaces so the biochemical details are still unknown. However, by modelling the interaction on other AAA+ proteins the association of neighbouring domains I and III is likely (Erzberger *et al.,* 2002). The combination of DnaA binding to *oriC*, DnaA oligomerisation and the resulting duplex bending causes a conformational change and local unwinding of the DUE (Messer, 2002). The remaining steps of initiation are summarised in Figure 1.2, steps C-E. Following step E the remaining components of the replisome can be loaded and elongation begins (see 1.6).

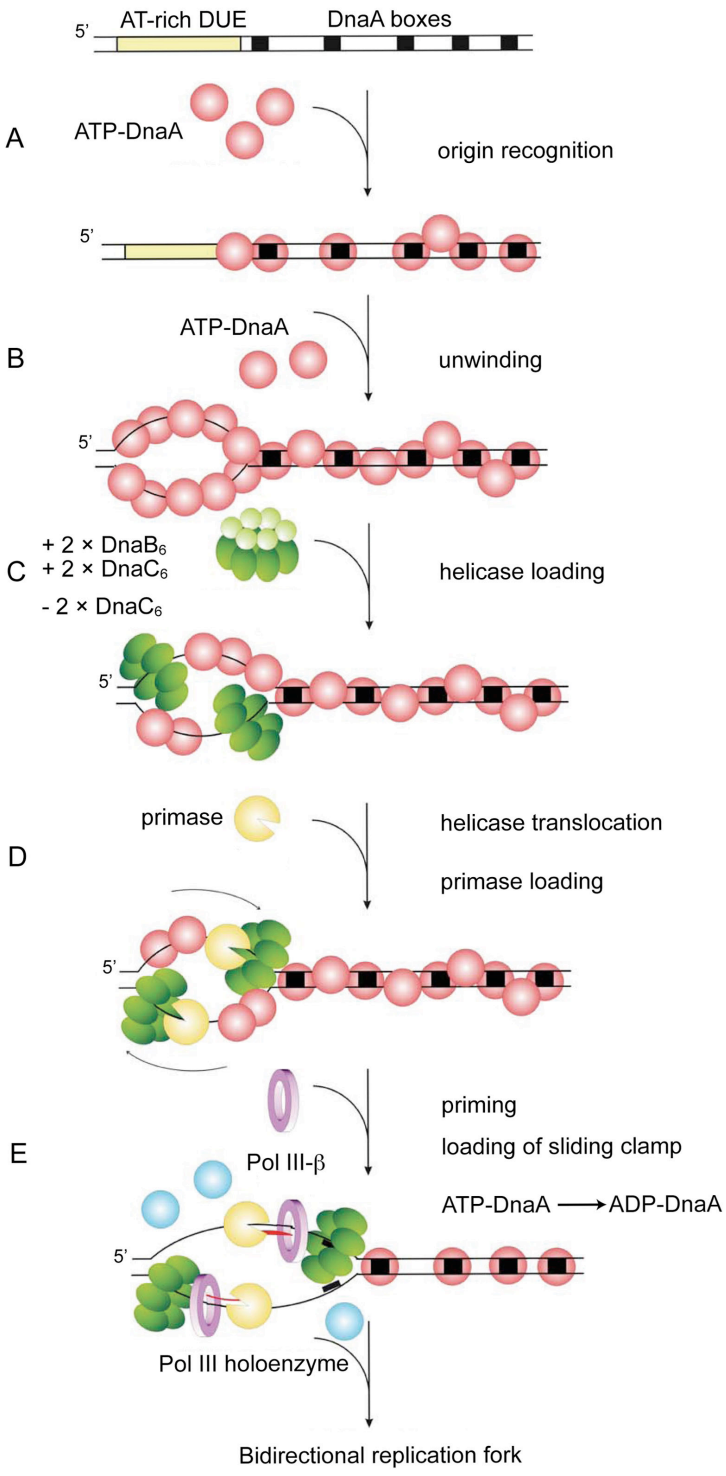


Figure 1.2: Replication initiation in *Escherichia coli*

Steps of replisome formation, adapted from Messer *et al.,* 2002.

(A) Recognition of *oriC* by DnaA binding to DnaA boxes.

(B) Distortion of duplex from DnaA binding and bound initiator protein oligomerisation causes DUE unwinding.

(C) DnaA positions two double hexamers of DnaB and DnaC on unwound DNA.

(D) DnaC leaves the complex and the accompanying ATP hydrolysis activates DnaB helicase activity. The DnaB hexamers move past each other (5′-3′) and expand the DNA bubble further so that DNA primase (DnaG) can enter.

(E) The clamp loader (Polymerase III γ complex) loads the sliding clamp (Polymerase III β-subunit) on to the primed template.

1.2.1 Bacterial replication origin regulation

There are many ways in which initiation is limited to a once per generation event. The ADP and ATP complexed forms of DnaA cycle but because only ATP-DnaA can induce DUE remodelling, the position of equilibrium is a target for regulation. Sliding clamp recruitment activates the intrinsic ATPase activity of DnaA (Katayama

et al., 1998) in conjunction with Hda (Kato and Katayama, 2001) and HU (Dixon and Kornberg, 1984). This prevents further initiation events by inactivating DnaA and is called regulatory inactivation. Newly replicated DNA is hemimethylated because only the parental strand has persisted long enough to be methylated by Dam methyltransferase. The eleven Dam methyltransferase recognition sequences (GATC) within *oriC* must be methylated for initiation of replication to occur (Smith *et al.,* 1985). These regions are sequestered from methylation by binding with other proteins such as SeqA (Lu *et al.,* 1994; von Freiesleben *et al.,* 1994), and SeqB (Shakibai *et al.,* 1998). SeqA can also reduce the amount of DnaA in a cell by preventing methylation of the *dnaA* promoter (Lu *et al.,* 1994), it may also have a role in inhibiting duplex unwinding by inducing positive supercoiling (Klungsoyr and Skarstad, 2004). DnaA acts as its own inhibitor (Braun *et al.,* 1985) and has recently been shown to contribute to its own sequestration (Bach *et al.,* 2008). Another mechanism to limit initiation is to channel DnaA to a "sink" in order to prevent it from binding *oriC*. The DnaA box motif is overrepresented in the *E. coli* genome (Mackiewicz *et al.,* 2004) and over 300 DnaA boxes can bind DnaA protein (Messer, 2002). Near *oriC* is a cluster of DnaA boxes called *datA*. This locus binds DnaA *in vitro* with a higher affinity than *oriC* (Kitagawa *et al.,* 1996), and additional *datA* loci have been shown to limit or block initiation (Kitagawa *et al.,* 1998). Therefore *datA* is thought to be a negative regulator of origin firing which functions by reducing the pool of initiator molecules. A combination of these regulatory pathways ensures that *oriC* does not over-initiate DNA replication.

In a study of 120 bacterial genomes Mackiewicz *et al* concluded that:

1) *dnaA* is ubiquitous amongst bacteria (except for the reduced genomes of endosymbionts *Wigglesworthia glossinidia* and *Blochmannia floridanus*).
2) The DnaA box is highly conserved between different bacterial species (with the exception of the proteobacteria).
3) Most genomes contain an overrepresentation of the DnaA box and many show DnaA box clustering. (Mackiewicz *et al.,* 2004)

This is good evidence that the replicon model adhered to by the DnaA-*oriC* system is widespread in bacteria. It also suggests that methods of initiation regulation are similar across the domain since the clusters of DnaA boxes may fulfil the role of *datA* in *E. coli*.

1.3 Eukaryotic DNA replication

The study of eukaryotic DNA replication origins has been complicated by the lack of a domain-wide paradigm. While bacteria use a DnaA-*oriC* system, eukaryotes replicate their genomes using a variety of strategies. Eukaryotes face additional challenges because their chromosomes are linear and larger than bacterial genomes. Multiple chromosomes are an additional consideration and pose a significant regulation problem. The question of how eukaryotes accomplish replication of the entire genome during S phase was answered before any eukaryotic origins were identified. In 1968 mammalian DNA synthesis was visualised using autoradiography and multiple replication forks were visible on an individual DNA strand (Huberman and Riggs, 1968; Taylor, 1968). The existence of multiple origins of replication on single replicons was a clear departure from the situation in bacteria. There were also early hints that eukaryotic origins were flexible. Conditional origins fired at different times depending on growth conditions and developmental stage, origin firing also appeared to be asynchronous (Gilbert, 2004 and references therein). The replicon model was formulated with bacteria in mind so there was no compelling reason for it to apply equally well to eukaryotes. Nevertheless, geneticists adhering to the replicon model began to search eukaryotic genomes for replicator sequences.

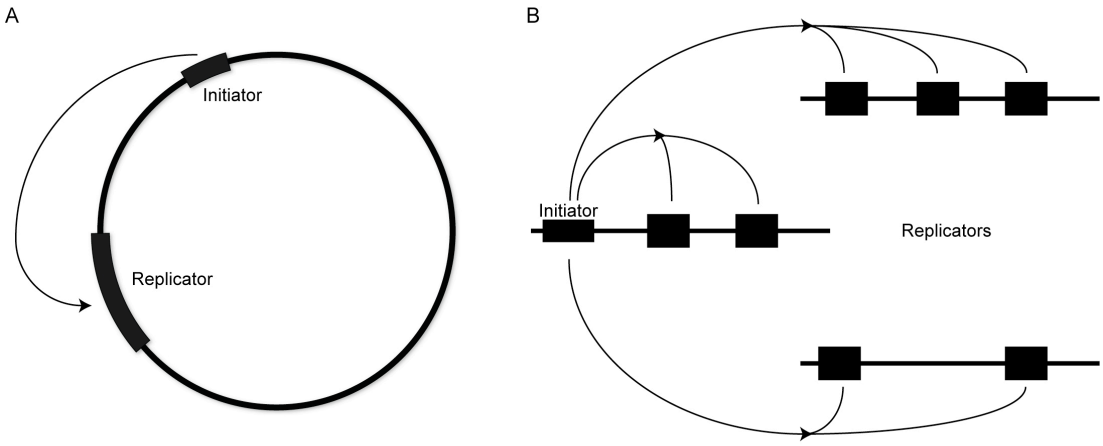


Figure 1.3: Replicon models
(A) Original replicon model applied to bacterial chromosome.
(B) Replicon model applied to eukaryotes with multiple, linear chromosomes.
Adapted from Gilbert, 2004.

1.3.1 Budding yeast

The unicellular budding yeast *Saccharomyces cerevisiae* is a common model eukaryote because it is easy to culture and manipulate. As yeast genetics advanced in the 1970s it was a natural candidate for the study of eukaryotic DNA replication. Transformation of *S. cerevisiae* with yeast vectors containing *S. cerevisiae* DNA fragments and a selectable marker resulted in different classes of transformants. While some plasmids integrated in to the yeast genome, others could replicate autonomously, were maintained episomally and had a much higher transformation efficiency (Struhl *et al.,* 1979). This group of sequences became known as Autonomously Replicating Sequences (ARS) (Stinchcomb *et al.,* 1979). Variations of this ARS assay have been used to identify origins of replication ever since. *S. cerevisiae* ARS elements are ~100 bp in length and are characterised by an essential consensus sequence: 5′-[A/T]TTTAT[A/G]TTT[A/T], called the A element (Kornberg, 1992). Many, but not all, *S. cerevisiae* ARS elements have been shown to function as origins of replication *in vivo* by visualising replication bubbles using two-dimensional gel electrophoresis (see p24) (Huberman *et al.,* 1987; Brewer and Fangman, 1987; Huberman *et al.,* 1988). The A element or ACS (ARS consensus sequence) is necessary but not sufficient for origin activity (Marahrens and Stillman, 1992) because the number of ACS matches in the genome vastly exceeds the number of functional replication origins *in vivo* (Nieduszynski *et al.,* 2006).

With the *S. cerevisiae* replicator sequence determined, the search began for the initiator protein. Using DNaseI foot printing at ARS1, the *S. cerevisiae* initiator was found to be a multiprotein complex called the Origin Recognition Complex (ORC) (Bell and Stillman, 1992). Mutations in the ACS or in genes encoding ORC subunits result in a loss of replicator activity and replication defects respectively. The ORC has been shown by several techniques to bind to *S. cerevisiae* origins of replication *in vivo* (Bell, 2002 and references therein). These findings support the initiator function of the ORC. Further molecular analysis of a specific origin (ARS1), revealed three more elements that are individually dispensable but collectively important for origin function B1, B2 and B3 (Marahrens and Stillman, 1992). The ACS and B1 elements each contain half of the ORC binding site (Rao and Stillman, 1995; Rowley *et al.,* 1995). The B elements are divergent and interchangeable so the multiple origins in *S. cerevisiae* are not identical. In addition to these sequence

elements, the 3′ flanking regions of the ACS at origins are A/T-rich and serve as the DUE (Newlon and Theis, 1993).

The ORC is analogous to DnaA, it binds sequence specifically to replicators, recruits replication factors and is a member of the AAA+ family of ATPases. The ORC is made of six proteins (Orc1p-Orc6p) and multiple ORC subunits interact with the A-rich strand of the ACS and the B1 element (Lee and Bell, 1997). Orc1p, Orc2p, Orc4p and Orc5p bind to DNA while Orc6p is not required for sequence-specific DNA binding. Orc3p does not bind directly but is still required which suggests it has a role in positioning the other ORC subunits (Lee and Bell, 1997). In the absence of a crystal structure, modelling based on a combination of imaging and binding assays has been used to predict how the ORC binds to DNA. The suggested model is that origin DNA runs along the length of the origin recognition complex with Orc1/4 close to the ACS and Orc2/3 near the B1 element (Chen *et al.,* 2008). The double binding site gives polarity to the interaction and the model predicts that this points the non-AAA+ domains of Orc6 and Orc2 towards the B2 element. Given that the number of ORC binding sites in the *S. cerevisiae* genome is less than the number of ACS matches (Wyrick *et al.,* 2001), it is likely that DNA sequence specificity alone is not sufficient to localise ORC to origins. There is evidence suggesting that other replication factors enhance ORC-ACS interactions, such as DNA-bound proteins recruiting ORC and a role for chromatin structure in regulating ORC distribution (Bell, 2002).

S. cerevisiae conforms to the replicon model, its replicator sequence (ACS) and initiator protein (ORC) can be readily identified. The key difference lies in the fact that budding yeast chromosomes contain multiple origins that vary in their molecular structure and the conditions under which they fire.

1.3.2 Fission yeast

Saccharomyces pombe replicators were also identified through ARS assays (Clyne and Kelly, 1995 and references therein). Unlike *S. cerevisiae*, these ARS elements are large (0.5-1 kb) and contain no well-defined consensus sequence (Cvetic and Walter, 2005). *S. pombe* ARS are A/T-rich asymmetric intergenic regions that show no direct similarities with each other. Some ARS contain essential regions but they

are not common between different origins and replacement with poly(dA/dT) tracts restores origin function (Okuno *et al.,* 1999). Homologues of the ORC are also found in the *S. pombe* genome (Gavin *et al.,* 1995). Given that A/T-content and tract length are the only determinants of origin activity in intergenic regions, it was proposed that *S. pombe* uses a sequence-independent DNA binding motif for ORC binding (Dai *et al.,* 2005). The *S. pombe* ORC structure goes some way to explaining the A/T-content dependence of fission yeast origins. The N-terminal DNA binding domain of the Orc4p subunit contains nine copies of an A/T-hook motif (Chuang and Kelly, 1999). These motifs bind DNA by "hooking" in to the minor groove of A/T tracts (Huth *et al.,* 1997). The A/T-hook motifs are essential and sufficient for ORC binding of ARS (Kong and DePamphilis, 2001; Lee *et al.,* 2001). As in *S. cerevisiae*, there are more potential ORC binding sites (*S. pombe* intergenic regions are ~70% A/T-rich) than active origins so again there are other factors which regulate potential origin firing.

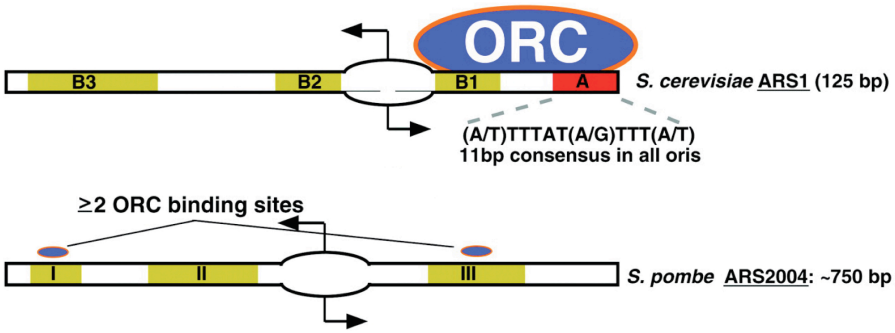


Figure 1.4: Replication origins in unicellular eukaryotes
Structure of replication origins in *S. cerevisiae* and *S. pombe*. Consensus ORC binding sites are indicated in red and sequences important for origin activity are shown in brown. Regions I, II and III are A/T-rich sequences. Replication origins are indicated by the bidirectional arrows passing through a bubble. Adapted from Gilbert, 2001.

Fission yeast does not conform to the replicon model as neatly as budding yeast. It has a clear initiator protein but the replicator sequences are defined by the broad characteristic of A/T-richness as opposed to sequence specificity. Despite being related yeast species, *S. cerevisiae* and *S. pombe* differ in their manner of origin determination. After a promising beginning the search for a universal eukaryotic replicator fell at the second hurdle. This proved to bode ill in the search for replicators in metazoa.

1.3.3 Metazoa

The ARS assay was the initial strategy used to identify replicators in metazoa. Despite the use of multiple model systems and the many variations of the method that were tried, an equivalent of the yeast ACS could not be identified (Gilbert, 2004). Using Epstein-Barr viral DNA to increase nuclear retention of plasmids improved ARS assay sensitivity and using this technique sequence independent replication was shown for human DNA on a plasmid within human cells (Krysan *et al.,* 1989; Krysan and Calos, 1991). The first hints of this "randomness" were actually noted in the early eighties when *Xenopus laevis* eggs were shown to support replication of viral, bacterial or eukaryotic DNA with or without known origins (Harland and Laskey, 1980; Mechali and Kearsey, 1984). These results were mostly classed as an exception to the replicon rule, partly because *X. laevis* are unusual in requiring entire genome replication within 20 minutes and partly due to belief in the replicon model (Gilbert, 2004). With the failure of the ARS assay to produce non-random results, alternative methods were developed to enable the study and mapping of replication origins. Unfortunately different methods often yielded contradictory results due to variation in their sensitivity. An example is the dihydrofolate reductase locus (DHFR) in *Cricetulus griseus* (Chinese hamster). Mapped by Okazaki fragment hybridisation to an array, >80% of initiation events occur at the same site (Burhans *et al.,* 1990). On the other hand initiation seemed to be random in the same region when it was mapped by two-dimensional agarose gel electrophoresis (Vaughn *et al.,* 1990). Gradually the study of metazoan origins has shown that they are much more complicated than the replicon model.

Metazoan origins can be broadly separated in to two classes. The human *β*-globin origin locus was identified using a replication direction assay; it can initiate from ectopic locations and was initially thought to contain a distinct origin of replication (Kitsberg *et al.,* 1993; Aladjem *et al.,* 1998). This was heralded as evidence that replicators existed and were necessary for replication in animal cells (Huberman, 1998). However, detailed mapping of this locus revealed that it was more complicated than initially thought because it actually contains two independent non-overlapping origins (Wang *et al.,* 2004). The human *β*-globin locus and the human lamin B2 locus are examples of metazoan origins that are localised to a few kilobases (Gilbert, 2001). In contrast, the *C. griseus* DHFR locus is a 55 kb intergenic region

that acts as an initiation zone with replication origins emanating from a multitude of locations (Vaughn *et al.,* 1990; Dijkwel and Hamlin, 1995; Sasaki *et al.,* 2006). Despite the marked origin flexibility in this region there are preferred initiation sites called *oriβ*, *oriγ* (Dijkwel and Hamlin, 1995), and *oriβ′* (Kobayashi *et al.,* 1998). Deletion analysis confirms that these preferred sites are not essential and surprisingly up to 70% of the DHFR intergenic region can be deleted without abolishing origin activity in the remaining region (Kalejta *et al.,* 1998). The DHFR locus is an example of the second type of metazoan origin which consists of multiple dispersed origins within an "initiation zone" (Gilbert, 2001). Within these classes there are no conserved replicators but intergenic regions are favoured and an A/T-rich DUE has been shown to be essential in many of the studied metazoan origins (DePamphilis, 1999).

As in yeasts the ORC is the metazoan initiator. The ORC is conserved throughout Eukarya and homologues have been found in every sequenced eukaryotic genome (Bell, 2002). The ORC acts as a "landing pad" for the replication machinery it recruits and in addition to its initiator role serves multiple alternative functions (reviewed in Sasaki and Gilbert, 2007). When classical methods to define metazoan replicators failed, many searched for origins by hunting for ORC binding sites *in vivo* (Gilbert, 2004). The ORC has been shown to bind the *Homo sapiens* lamin B2 origin (Cvetic and Walter, 2005) and the *Drosophila melanogaster* chorion origin locus on chromosome three (Austin *et al.,* 1999). *D. melanogaster* ORC can preferentially bind negatively supercoiled DNA (Remus *et al.,* 2004) but in keeping with the lack of a replicator the metazoan ORC shows no sequence specificity.

In the absence of a sequence-specific replicator-initiator interaction, a bewildering array of factors that regulate origins in metazoa have been reported. Disruption of DHFR transcription reduces initiation events in the intergenic region and stimulates abnormal origin firing within the DHFR gene (Saha *et al.,* 2004). Early replicating *D. melanogaster* regions correlate with transcriptionally active locations (MacAlpine *et al.,* 2004) and plasmids injected in to *X. laevis* eggs show a preference for initiating near strong promoters (Danis *et al.,* 2004). In both these species the switching on of transcription at the mid-blastula transition causes initiation to change from a random process to a more specific event limited to intergenic regions (Hyrien *et al.,* 1995; Sasaki *et al.,* 1999). Transcription is also relevant in *S. cerevisiae* where

there is a hierarchy of preferential initiation of ARS which correlates with local transcription patterns (Donato *et al.,* 2006). These examples and the earlier point that origins tend to be in intergenic regions illustrate the importance of transcriptional interference with origin activity. Deletion of preferred origins within the DHFR intergenic region without changing the overall replication efficiency of the locus shows that normally suppressed alternative origins are firing (Kalejta *et al.,* 1998). The inhibition of an origin by another is known as origin interference (Brewer and Fangman, 1993) and may be a method of regulating broad initiation zones. Deletion of distal sequences 50 kb upstream of the human *β*-globin locus results in a loss of origin function (Aladjem *et al.,* 1995). This implicates distant DNA sequences in initiation site determination roles. Methylation and acetylation have been shown to act as inhibitor and promoters of replication respectively while chromatin structure can also influence origin choice (reviewed in Cvetic and Walter, 2005). The interaction of these different modes of regulation gives eukaryotic origins extensive flexibility. In bacteria regulation is achieved by preventing initiator-replicator binding. In eukaryotes the multiple factors that influence ORC binding to potential origins are extra layers of regulation and can be manipulated to control origin efficiency and timing to ensure that over-replication is avoided.

The lack of a conserved replicator in metazoa demonstrates that the replicon model is not applicable to eukaryotes other than *S. cerevisiae*. The desperate search for a conserved sequence analogous to *oriC* and the ACS has gradually been replaced with a desire to understand the ubiquity of the ORC and the divergent nature of eukaryotic origins and their regulation (see Figure 1.5).

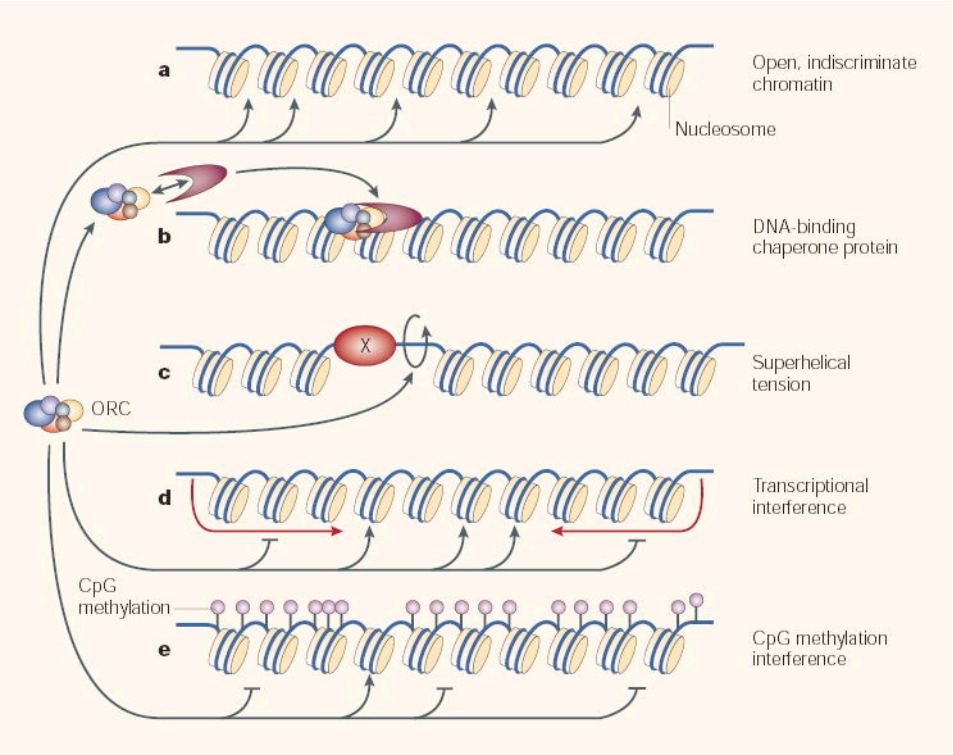


Figure 1.5: Relaxed replicon model
This revised replicon model reflects the context-dependent scheme of selecting metazoan origins. Factors which influence ORC binding are shown. From Gilbert, 2004.

1.3.4 Replisome formation in eukaryotes

Despite metazoan origins not being as defined as those in *S. cerevisiae*, most of the yeast replication factors are highly conserved amongst eukaryotes (Bell and Dutta, 2002). This implies that the events leading to origin processing are also conserved. In contrast to bacteria the initiator-origin binding is not sufficient for replication to begin. In metazoa the ORC remains bound to DNA throughout most of the cell cycle so an intermediate licensing step is required for origin firing (Diffley *et al.,* 1994).

Given the functional similarity between DnaA and the ORC, identification of a eukaryotic helicase that acted at replication forks and in initiation like DnaB was expected. Despite the multiple putative helicases encoded by eukaryotic genomes, the identity of the replicative helicase has yet to be proven. The prime candidate is the minichromosome maintenance complex (MCM) which is a hexamer of Mcm2-7p (Tye, 1999). Unfortunately no hexamer helicase activity has been shown *in vitro* for this eukaryotic complex. However, helicase activity has been shown *in vitro* for the

human Mcm4/6/7 sub-complex (Ishimi, 1997) and there are tantalising hints of a role for Mcm2-7p in initiation and elongation from *in vivo* and *in vitro* work in other eukaryotes (reviewed in Labib and Diffley, 2001) and archaea (see 1.5.3). Problems demonstrating *in vitro* helicase processivity may stem from the use of artificial substrates that might not sufficiently resemble eukaryotic replication forks (Sclafani and Holzen, 2007). It is also possible that these reactions are missing replication factors essential to MCM helicase activity. Despite these problems the MCM complex is generally assumed to be the replicative helicase and there are several models that describe its unwinding mechanism (reviewed in Sclafani and Holzen, 2007).

Origin bound ORC recruits the replication factors Cdc6p (Cdc18p in *S. pombe*) and Cdt1p. Cdc6p is an AAA+ ATPase with sequence similarity to Orc1p and it functions as a helicase loader (Bell *et al.,* 1995; Neuwald *et al.,* 1999). When the ORC is bound to origin DNA it can bind Cdc6p *in vitro* (Mizushima *et al.,* 2000) and Cdc6 ATPase activity has been shown to regulate the ORC-Cdc6 complex stability *in vitro* (Speck and Stillman, 2007). Both Cdc6 and Cdt1 are essential for helicase loading. The pre-replicative complex formation is complete once the Mcm2-7 complex is loaded and the origin is then licensed for replication. However, induction of replisome formation only occurs once the helicase has been activated. Helicase activation requires phosphorylation by cyclin-dependent kinases (CDKs) and Dbf4-dependent kinases (DDKs) in conjunction with the binding of additional replication factors. Dpb11, Sld3, Cdc45 and the GINS complex are replication proteins which interact and assemble at replication origins and are important for replisome formation (Takayama *et al.,* 2003). For example, it is thought that the GINS complex couples the MCM helicase to replication factors at replication forks (Labib and Gambus, 2007). DNA synthesis initiation is achieved by the recruitment of replisome components and their subsequent functional interactions. The eukaryotic replisome proteins are unrelated to those in bacteria but the same functional divisions are apparent (see Figure 1.20).

As in bacteria, initiation regulation prevents over-replication of the genome. Eukaryotes only replicate DNA during the S-phase of the cell cycle, to achieve this there are multiple levels of regulation linked to the cell cycle. CDKs are important in three inhibitory pathways: phosphorylation of the ORC, down-regulation of Cdc6

activity and nuclear exclusion of the MCM complex (Nguyen *et al.,* 2001). MCM complex loading is blocked during S, G2 and M phase to inhibit the licensing step. In conjunction with nuclear exclusion of the MCM complex, this is achieved in metazoa by geminin binding which inhibits the helicase loader Cdt1. Helicase activators are also regulation targets. CDKs and DDKs are activated by the binding of their unstable subunit (cyclin and Dbf4 respectively). Cell cycle regulation of these subunits is also used as a means of origin regulation. For example, Dbf4 is degraded during G1 and only during S-phase does CDK phosphorylation inhibit the degradation machinery so that Dbf4 levels are sufficient to activate DDKs (Sclafani and Holzen, 2007 and references therein).

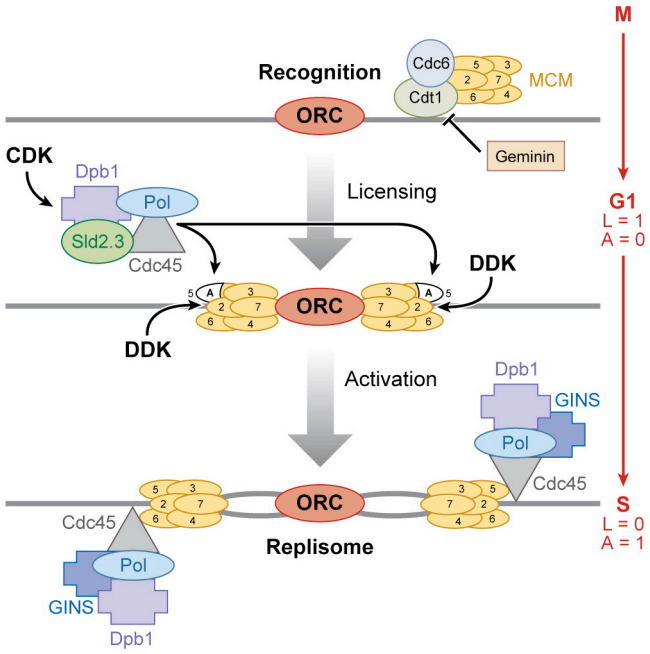


Figure 1.6: Eukaryotic replication initiation regulation
Eukaryotic origins are recognized by the ORC, then Cdc6 and Cdt1 protein load the hexameric MCM helicase to form the "licensed" (L) pre-RC in G1 phase (L = 1, A = 0). CDK and DDK are active in late G1 and activate (A) the MCM helicase so that the replisome can be loaded. CDK inhibits any further licensing (L = 0, A = 1). From Sclafani and Holzen, 2007.

Many details of replisome formation differ depending on the system. For example, *S. cerevisiae* has one CDK while mammals have many and the time of action relative to DDK differs between *S. cerevisiae* and *X. laevis* (Sclafani and Holzen, 2007). Despite the differences in the details, the general mode of replisome formation and recruited replication factors after ORC binding is similar across eukaryotes.

1.4 Archaea

After describing replicators and initiators in Bacteria and Eukarya the third domain of life will now be considered. Until the advent of molecular biology, organisms were categorised primarily based on morphology and physiology. Unfortunately there are inherent difficulties to this approach when constructing phylogenies. Morphology can be misleading when trying to establish relationships between species demonstrating convergent evolution. Another difficulty is that microbial physiology is often difficult to interpret. To add to the limited number of classical phenotypes and to aid phylogeny construction it was suggested that the informational content of macromolecules could be mined (Zuckerkandl and Pauling, 1965). These authors realised that comparisons of homologous molecules from different species could reveal "the approximate time of existence of a molecular ancestor common to the chains that are being compared". Such comparisons could therefore provide an index of species relationships and they have become the dominant method of phylogeny construction.

The dominant view of life in the 20th century was seen through the window of the prokaryotic-eukaryotic divide. This classifies Eukarya as species with a nucleus enclosed by a nuclear membrane and Prokarya as those without a true nucleus. However, a group of prokaryotes with unique properties like methanogenesis were the first hint that the prokaryotic world was more diverse than originally thought (for example see (Zeikus and Wolfe, 1972)). Once DNA sequencing methods were feasible it was possible for DNA to form the basis of taxonomy rather than protein molecules. Carl Woese and George Fox used this to great effect by taking 16S (small-subunit) ribosomal RNA (rRNA) nucleotide sequences as the basis for a phylogenetic tree. rRNA was a good choice because it is universal and contains different domains evolving at different speeds. This means that rRNA can be used for fine and large-scale measurements of evolutionary distance (Woese, 1987). The results suggested that the unusual methanogens were actually distinct from bacteria. Woese and Fox named them "archaebacteria" and proposed that established phylogeny be reorganised to reflect a tripartite division of life (Woese and Fox, 1977). As more archaebacteria were isolated and their distinct properties were characterised it became clear that they were a monophyletic group (Fox *et al.,* 1980).

Contrary to expectations archaebacteria were more closely related to eukaryotes than bacteria. In 1990 it was proposed that archaebacteria should be renamed "archaea" to avoid the implication that they were closely related to bacteria (Woese *et al.,* 1990). These authors also proposed that Archaea, Eukarya and Bacteria should be regarded as the highest level of classification and the name "domain" was suggested for this new taxon. The three domain concept has gradually been accepted (Cavicchioli, 2007) despite initial resistance and the remaining doubters (Mayr, 1998). Contrary to calls for the prokaryotic-eukaryotic divide to be abandoned on the grounds that prokaryote is a meaningless negative definition and not a monophyletic group; most "trees of life" group Archaea and Bacteria together as prokaryotes.

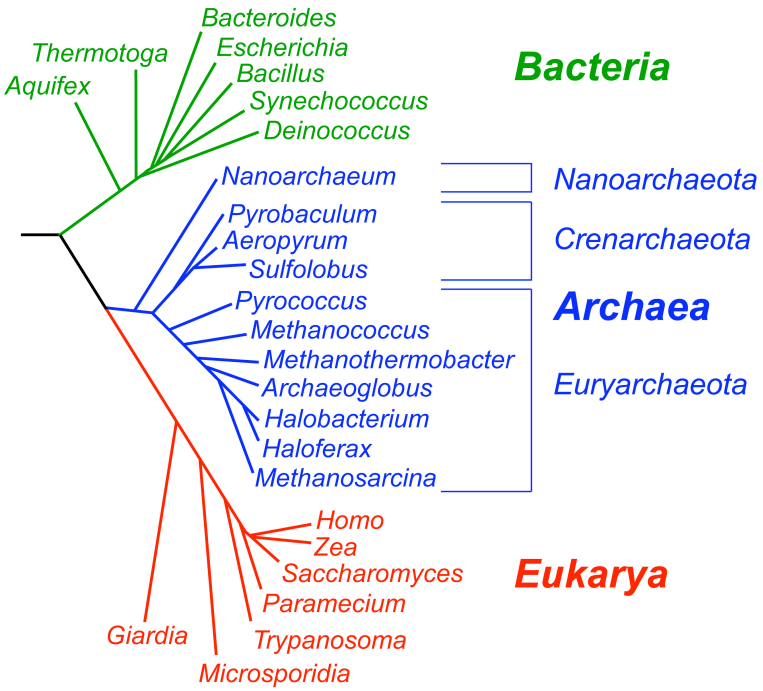


Figure 1.7: The "tree of life"
Phylogenetic tree based on rRNA sequences. The three domains of life and selected phyla and genera are shown. From Allers and Mevarech, 2005.

The Archaea are split in to two main phyla: the diverse Euryarchaeota and the mostly hyperthermophilic Crenarchaeota (Woese *et al.,* 1990). Based on the identification of new species and genome sequencing, the phyla Nanoarchaeota (Huber *et al.,* 2002), Korarchaeota (Elkins *et al.,* 2008) and Thaumarchaeota (Brochier-Armanet *et al.,* 2008) have recently been characterised. Due to the limited examples of these groups, the evolutionary relationships within the Archaeal domain remain unclear.

1.4.1 Characteristics of the Archaea

Archaea are well known for their extremophile lifestyles. Different representatives of the domain can thrive despite extremes of temperature, pH and salinity (Allers and Mevarech, 2005). However, the relevance of mesophilic archaea has been underscored by the estimate that archaea might comprise 20% of total world biomass (DeLong and Pace, 2001) and are much more widespread than previously thought (Venter *et al.,* 2004). Befitting of their status as the third domain of life, archaea contain a mix of unique characteristics in addition to those which are related to bacterial or eukaryotic features. Even before the first archaeal genome was sequenced they were described as having a "eubacterial form with eukaryotic content" (Keeling *et al.,* 1994). In general, information-processing genes (replication, translation, transcription) are more similar to those in eukaryotes while housekeeping genes (core metabolism) resemble those found in bacteria, but this is not an absolute division. For example, the archaeal RNA polymerase core and the basal transcription machinery is more similar to the eukaryotic version but in contrast many archaea have homologues to bacterial transcriptional regulators (Aravind and Koonin, 1999; Werner, 2007). The archaeal translation machinery is eukaryotic in nature (Londei, 2005). Though archaea use polycistronic mRNAs they use a similar number of translation initiation factors to eukaryotes (>10 as opposed to 3 in bacteria) (Dennis, 1997). Archaeal genomes are structured like bacteria with circular chromosomes and their morphology is bacteria-like. However, euryarchaeota contain histones and therefore may package DNA in a similar way to eukaryotes (White and Bell, 2002).

When the first archaeon was sequenced, 56% of its open reading frames showed no homology with previously annotated genes (ORFans) (Bult *et al.,* 1996). As more genomes have become available for comparison the number of ORFans has been reduced but there are still a significant number (10-20%) of archaeal genes with unknown functions (Makarova *et al.,* 2007; Koonin and Wolf, 2008). This suggests a rich array of novel pathways and proteins unique to the archaea. A known example is methanogenesis, a metabolic strategy unique to the third domain of life. Archaeal cell membranes consist of a rigid monolayer with isoprene side chains that are ether-linked to glycerol. This structure makes them relatively impermeable and therefore suitable for extreme environments (van de Vossenberg *et al.,* 1998). Many proteins

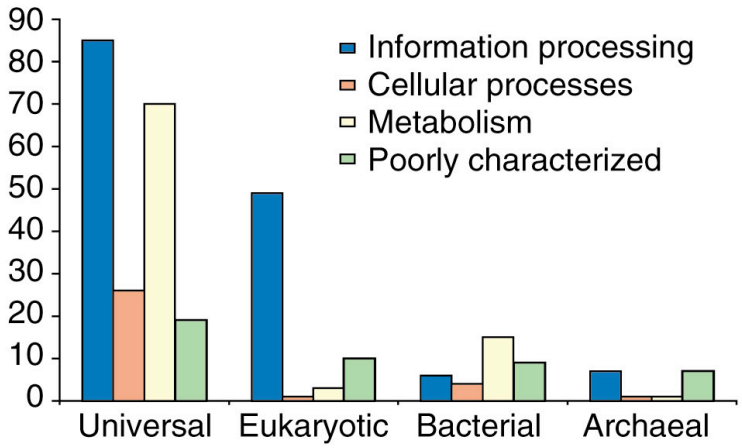


Figure 1.8: Archaeal genes
Functional breakdown of genes within the conserved archaeal core, showing the abundance of eukaryotic-like information processing genes in archaea. The data for orthology and functional classification are derived from the clusters of orthologous groups. From Makarova and Koonin, 2003.

will be unique because they have evolved to tolerate the implications of their extreme environments. Practical uses have been found for archaeal molecules specialised for functioning in extreme environments e.g. PCR polymerases and biocatalysts (Egorova and Antranikian, 2007).

While the Archaea are interesting in their own right, they are often studied as models of eukaryotes, particularly for DNA replication and repair. Structural biology has also benefited from the increasing amount of archaeal research. The extremophilic nature of many studied species mean that their proteins have properties which make them amenable to crystallisation. For example, the first structural information on ribosomes came from a subunit purified from the halophilic archaeon *Haloarcula marismortui* (Ban *et al.,* 2000). Archaeal protein crystal structures are often used as a basis for the extrapolation of ideas concerning the eukaryotic versions. In this field great contributions have been made from ORC structural studies and MCM functional assays in archaea. Genetic analysis has lagged behind but the recently developed ability to genetically manipulate various archaeal species has finally enabled genetic analysis of the third domain of life (reviewed in Allers and Mevarech, 2005; Rother and Metcalf, 2005).

1.5 Archaeal DNA replication

Archaea have small, circular genomes so it was expected that they would replicate their DNA in a bacterial manner; from a single origin and using replication factors related to those found in bacteria (Edgell and Doolittle, 1997). However, even before the first archaeal replication origins were characterised there were several signs that this was not the case. Halophilic archaea were shown to be sensitive to the eukaryotic DNA polymerase inhibitor aphidicolin (Forterre *et al.,* 1984). In addition, eukaryote-like B family DNA polymerases were identified in various euryarchaea and crenarchaea throughout the eighties and nineties (Edgell *et al.,* 1997 and references therein). Possession of eukaryotic replication machinery by an archaeon was confirmed when the first archaeal genome was sequenced (Bult *et al.,* 1996). The original analysis of the *Methanococcus jannaschii* (formerly *Methanocaldococcus jannaschii*) genome showed that it contained eukaryotic-like homologues for the replication processivity factor PCNA (proliferating cell nuclear antigen), the clamp loading complex (replication factor C), DNA polymerase, DNA ligase and Fen1 (primer removal) (reviewed in Edgell and Doolittle, 1997). Many archaeal genome sequencing projects were then completed in quick succession which rapidly increased the number of known archaeal replication proteins and enabled a more thorough comparison with those found in bacteria and eukaryotes (Bohlke *et al.,* 2002).

Putative archaeal replication proteins are all either homologues of the eukaryotic protein or show greater similarity to the eukaryotic rather than the bacterial version (Barry and Bell, 2006). The archaeal replication machinery can be said to represent a "streamlined" version of the eukaryotic system because many archaeal replication factors contain fewer subunits than those in eukaryotes (see Figure 1.20). For example, eukaryotes require a family of MCM proteins and a five subunit replication factor C while *Methanobacterium thermoautotrophicum* possesses only one *mcm* homologue and a replication factor C made of two subunits (Vas and Leatherwood, 2000).

The identification of ORC homologues within the Archaea was particularly intriguing. Almost all sequenced archaeal genomes contain at least one gene with equal similarity to *cdc6* and *orc1* (Barry and Bell, 2006). In eukaryotes, Cdc6 is

involved in loading the MCM complex at replication origins and Orc1 is one of the six subunits of the eukaryotic initiator protein. No convincing homologues of DnaA have been found amongst the archaea. Cdc6/Orc1 may function as the archaeal initiator and/or the helicase loader (see section 1.5.2).

With the archaeal replication protein set coming in to focus the pressing question became: "how do archaea replicate their bacteria-like chromosomes with eukaryotic-like replication machinery?". Genetic techniques lagged behind the wealth of sequence information from archaeal genomes so the first attempts to locate archaeal origins used bioinformatics. DNA polymerases are unidirectional and synthesise DNA from 5′ to 3′. Since replication forks are bidirectional one strand is replicated discontinuously and the other continuously. This results in different selection pressures on different DNA strands and therefore asymmetry in genomes. There are many different explanations for the existence of these patterns and it is likely that many contribute to the generation of asymmetry (reviewed in Frank and Lobry, 1999). The current consensus explanation is that to avoid C to T mutations as a result of single-strand cytosine deamination, there is a lack of cytosine in the leading strand which exists in a single-stranded form for longer during DNA replication (Sernova and Gelfand, 2008). The statistical measure of this bias is called a GC skew. GC skews change sign abruptly at replication origins because at these points strand synthesis switches between continuous and discontinuous modes of replication. There are several methods of analysing nucleotide and oligonucleotide compositional biases (reviewed in Sernova and Gelfand, 2008).

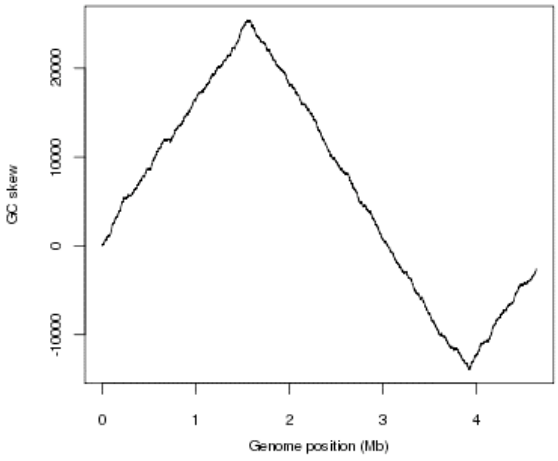


Figure 1.9: Example GC disparity curve for *Escherichia coli*
GC nucleotide skew for *Escherichia coli*. Skew inflection points represent the origin and terminus of replication. From Roten *et al.,* 2002.

Putative origins of replication were identified using GC/oligomer skew analysis for several archaeal species [*Archaeoglobus fulgidus* and *M. jannaschii*, (Grigoriev, 1998); *M. thermoautotrophicum* and *Pyrococcus horikoshii* (Lopez *et al.,* 1999)]. These *in silico* studies identified a single candidate origin for *M. thermoautotrophicum* and *P. horikoshii* but presented an unclear picture in *A. fulgidus* and *M. jannaschii*. The *A. fulgidus* skew plot showed two global maxima and minima, suggesting that this species might use two origins of replication. There was no identifiable consensus sequence equivalent to the ACS at these candidate origins but one common feature was their proximity to *cdc6*/*orc1* homologues. This is reminiscent of the situation in bacteria where *oriC* is located near, and often adjacent to *dnaA*. Another feature with similarity to bacterial origins was the presence of 13 bp repeats distributed around an A/T-rich element (Lopez *et al.,* 1999). These authors suggested that the repeats formed the recognition sequence for the Cdc6/Orc1 initiator in a similar manner to DnaA boxes.

Methods used to locate replication origins in archaea

Bioinformatic evidence needs to be supported by experimental confirmation of origin activity. The study of bacterial and eukaryotic origins produced a variety of techniques that could be used to characterise origins of replication in archaea. The principles behind three common methods are briefly described below.

Two-dimensional agarose gel electrophoresis

This technique relies upon two perpendicular electrophoretic separations of DNA molecules and exploits the fact that replication intermediates such as forks, replication bubbles and termination structures migrate to characteristic positions (Bell and Byers, 1983). The first dimension separates DNA based on size, and the second according to size and topology. Replication bubbles can be detected in individual restriction fragments using Southern hybridisation. This technique was developed to visualise recombination intermediates (Bell and Byers, 1979), but has since been widely used to confirm *in vivo* origin function (reviewed in Dijkwel and Hamlin, 1997).

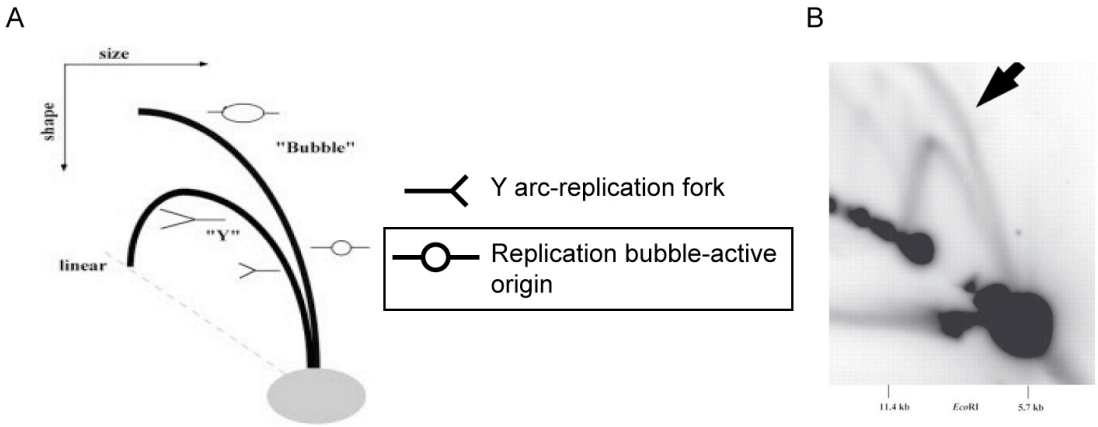


Figure 1.10: Two-dimensional agarose gel electrophoresis
(A) Diagram of replication intermediates after two-dimensional agarose gel electrophoresis and with characteristic migration positions of each species.
(B) Example of two-dimensional agarose gel electrophoresis for the replication origin of *Pyrococcus abyssi*. The bubble arc identified by the arrow demonstrates that a replication origin exists within this *Eco*RI fragment. From Matsunaga *et al.,* 2001.

RIP mapping

DNA replication is bidirectional and semi-discontinuous so there is a transition point between discontinuous and continuous synthesis at origins (see Figure 1.11). Replication initiation point mapping (RIP) locates this transition point using primer extension reactions which extend from a labelled primer to the DNA/RNA junctions of nascent strand template (see Figure 1.11) (Bielinsky and Gerbi, 1998). Reaction products can then be separated by electrophoresis to identify the transition point.

Marker frequency analysis

DNA sequences close to replication origins will be replicated earlier than more distant sequences. Frequency analysis of DNA from synchronised cells can be used to discern origin location because sequences close to replication origins will be enriched. Enrichment profiles can be obtained using DNA-DNA hybridisation to markers distributed across the region of interest (see Figure 1.12). This method typically uses microarrays.

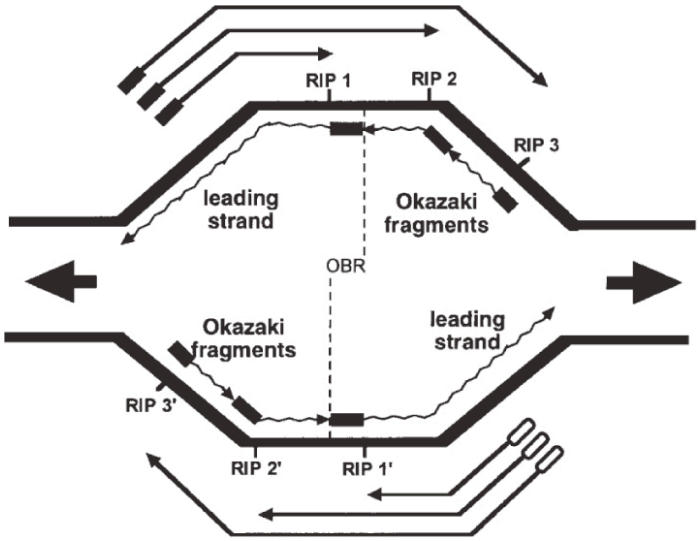


Figure 1.11: RIP mapping principle

Primer extension products (shown as arrows outside the replication bubble) stop at the DNA/RNA junction (RIP1, RIP2 etc). The transition point indicates the origin of bidirectional replication (OBR). From Gerbi and Bielinsky, 1997.

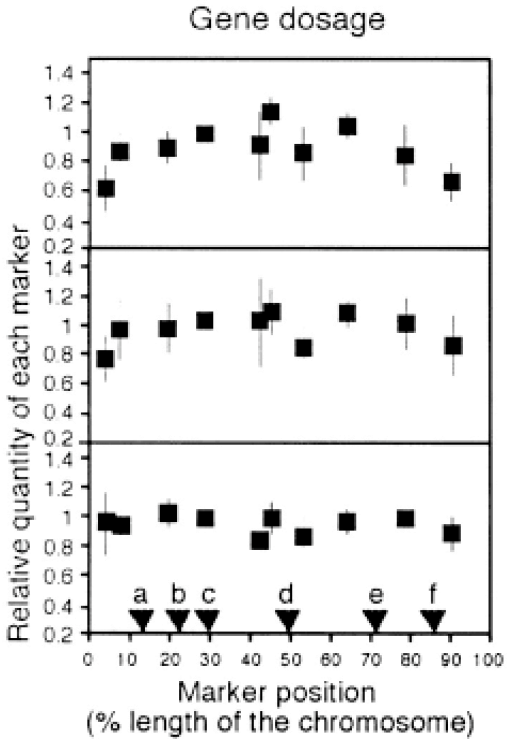


Figure 1.12: Example of marker frequency analysis

A. fulgidus strain VC-16 marker frequency analysis in exponentially growing cells at two time points and in stationary phase cells (bottom sample). Letters indicate the positions of different replication genes. The earliest time point (top panel) shows two marker abundance gradients from 45% to 5% and from 65% to 90% of the length of the chromosome. This region is therefore likely to contain the origin of replication. From Maisnier-Patin *et al.,* 2002.

The first archaeal origin to be characterised experimentally was in *Pyrococcus abyssi* (Myllykallio *et al.,* 2000). By combining nucleotide skew data, comparative genomics and *in vivo* labelling of early replicated DNA; these authors showed that *P. abyssi* has a single bidirectional origin of replication (*oriC*). In common with most of the putative origins identified *in silico*, the *P. abyssi oriC* is an intergenic region adjacent to a *cdc6/orc1* homologue. The *P. abyssi oriC* locus is near genes encoding a DNA polymerase, replication factor C (a polymerase accessory protein) and a helicase (Myllykallio *et al.,* 2000). The proximity of replication genes to an origin is similar to that seen in bacteria. The *in vivo* activity of this origin was later demonstrated using two-dimensional agarose gel electrophoresis (Matsunaga *et al.,* 2001), and the precise initiation point was determined by RIP mapping (Matsunaga *et al.,* 2003).

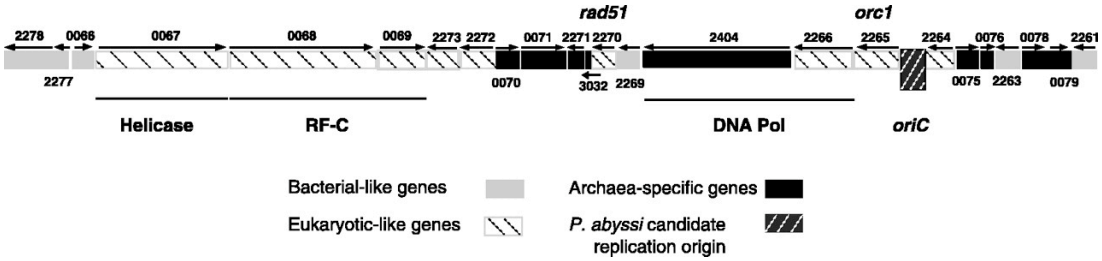


Figure 1.13: *Pyrococcus abyssi* replication origin
P. abyssi oriC locus and surrounding genes involved in DNA replication. From Myllykallio *et al.,* 2000.

The possibility of eukaryotic replication machinery replicating a circular chromosome with a single origin generated much excitement and speculation over the implications for the evolutionary relationships between the three domains. However, one data point makes for bad statistics and the picture was about to get more complicated.

GC skew analysis of *Halobacterium* sp. NRC-1 pointed towards two origins of replication on its main chromosome (Kennedy *et al.,* 2001). Z-curves represent a DNA sequence in three-dimensional space and can be used in a similar way to nucleotide skews to predict replication origin locations (Zhang and Zhang, 1994). Z-curve analysis of *Halobacterium* sp. NRC-1 also predicted two origins of replication on its main chromosome, both adjacent to *cdc6/orc1* homologues (Zhang and Zhang, 2003). Most archaea have one or two *cdc6/orc1* homologues but halophiles often

have many more (Barry and Bell, 2006), *Halobacterium* sp. NRC-1 is a good example with nine distributed throughout its genome. Zhang and Zhang also suggested three origin locations adjacent to *cdc6/orc1* homologues for *Sulfolobus solfataricus* (Zhang and Zhang, 2003).

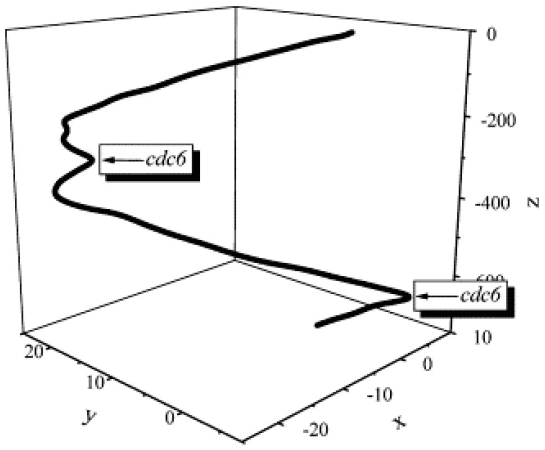


Figure 1.14: *Halobacterium* sp. NRC-1 Z-curve analysis
Z-curve analysis of the main *Halobacterium* sp. NRC-1 chromosome indicating two candidate replication origins adjacent to *cdc6/orc1* homologues. From Zhang and Zhang, 2003.

Verification of these predictions by experimental methods was only partially successful. Only one of the predicted origins for *Halobacterium* sp. NRC-1 was shown to function in ARS assays (Berquist and DasSarma, 2003). Single replication origins have also been identified for each of its two minichromosomes (Ng and DasSarma, 1993; Berquist and DasSarma, 2003). Skew analysis can fail to identify origins because of recent genome rearrangements or due to unique features of a particular species. In the case of *S. solfataricus* the origins adjacent to *cdc6-1* and *cdc6-3* were confirmed to function *in vivo* by two-dimensional agarose gel electrophoresis and RIP mapping (Robinson *et al.,* 2004). No origin activity could be found using these techniques in the vicinity of *cdc6-2* but at a later date a third origin was identified by marker frequency analysis. This origin was 50 kb distant from its predicted location near *cdc6-2* which showed that archaeal origins are not always directly associated with *cdc6/orc1* genes (Lundgren *et al.,* 2004). *S. solfataricus* was the first example of multiple origins of replication located on a single circular archaeal replicon.

Several more predicted archaeal origins have since been shown to function and are detailed in Table 1.1.

Table 1.1: Experimentally confirmed archaeal replication origins.
Genome size and replicon number from The UCSC Archaeal Genome Browser (http://archaea.ucsc.edu).

Species	Phyla	Genome size (Mb)	Replicon number	Origins	*cdc6/orc1*	Experimental confirmation	Reference(s)
P. abyssi	E	1.77	1	1	1	2-D-gel RIP mapping	(Myllykallio *et al.,* 2000) (Matsunaga *et al.,* 2003)
Archaeoglobus fulgidus	E	2.18	1	1	2[a]	MFA	(Maisnier-Patin *et al.,* 2002)
Halobacterium sp. NRC-1	E	2.57	3	1[b]	9	ARS assay	(Berquist and DasSarma, 2003)
S. solfataricus	C	2.99	1	3	3[c]	2-D-gel RIP mapping MFA	(Robinson *et al.,* 2004) (Lundgren *et al.,* 2004) (Robinson *et al.*, 2007)
Sulfolobus acidocaldarius	C	2.23	1	3	3[c]	MFA 2-D-gel	(Lundgren *et al.,* 2004) (Duggin *et al.,* 2008a)
Aeropyrum pernix	C	1.67	1	2	2	RIP mapping 2-D-gel	(Grainge *et al.,* 2006) (Robinson and Bell, 2007)
M. thermautotrophicus	E	1.75	1	1	2	2-D-gel	(Majernik and Chong, 2008)

E = Euryarchaeota C = Crenarchaeota MFA = marker frequency analysis

2-D-gel = two-dimensional gel electrophoresis

[a] Neither associated with the origin.

[b] The main chromosome contains a single origin. The two large extrachromosomal replicons, pNRC200 (365 kb) and pNRC100 (191 kb) contain a *repH*-dependent origin located within a common region (Ng and DasSarma, 1993).

[c] *cdc6-2* is not associated with an origin.

Fine mapping of these origins and comparisons with sequenced genomes of closely related species has meant that archaeal replication origins can now be described. The typical features are:

1. A/T-rich duplex unwinding elements
2. Origin recognition boxes
3. Associated with a *cdc6/orc1* homologue

1.5.1 Archaeal replicators

Archaeal origin loci contain short regions enriched for A/T content. For example, the origin of *Halobacterium* sp. NRC-1 contains a 189 bp region with an A/T content of 44% in comparison with a 32% genome average (Berquist and DasSarma, 2003). As in the other domains of life these A/T-rich regions are called duplex unwinding elements (DUE) because they are intrinsically meltable regions of DNA. Many DUE contain the predicted sites for DNA unwinding at the start of initiation. This is experimentally supported by the DUE at *oriC2* in *S. solfataricus* (79% A/T-rich), which contains the transition point between lagging and leading strand synthesis that marks the precise replication initiation site (Robinson *et al.,* 2004).

The multiple short inverted repeats noticed at predicted origins are similar in sequence and length across archaea. DNaseI footprinting assays showed that purified *S. solfataricus* Cdc6-1 bound these conserved sequences which were named origin recognition boxes (ORB) (Robinson *et al.,* 2004). These matched the inverted repeats noted previously at the *P. abyssi oriC* (Lopez *et al.,* 1999; Matsunaga *et al.,* 2003) and were also found at the functional *Halobacterium* sp. NRC-1 origin. ORB elements are conserved across the archaea (see Figure 1.15) and given that *S. solfataricus* Cdc6-1 can bind *P. abyssi* and *Halobacterium* sp. NRC-1 ORB elements, they also show functional conservation (Robinson *et al.,* 2004). Also characterised in the study by Robinson *et al*., is study was the mini-ORB, a sequence repeat related to the ORB core sequence and found without the full length ORB at *S. solfataricus oriC2*. The mini-ORB sequence is also broadly conserved across the archaea as shown in Figure 1.15. ORB elements with different orientations are found distributed (sometimes evenly) throughout intergenic origin loci and have even been noted within nearby genes (Matsunaga *et al.,* 2007). Pairs of inverted ORBs either

side of a DUE often exist at replication origins. The existence of multiple conserved repeats resembles the bacterial *oriC* which contains multiple DnaA boxes.

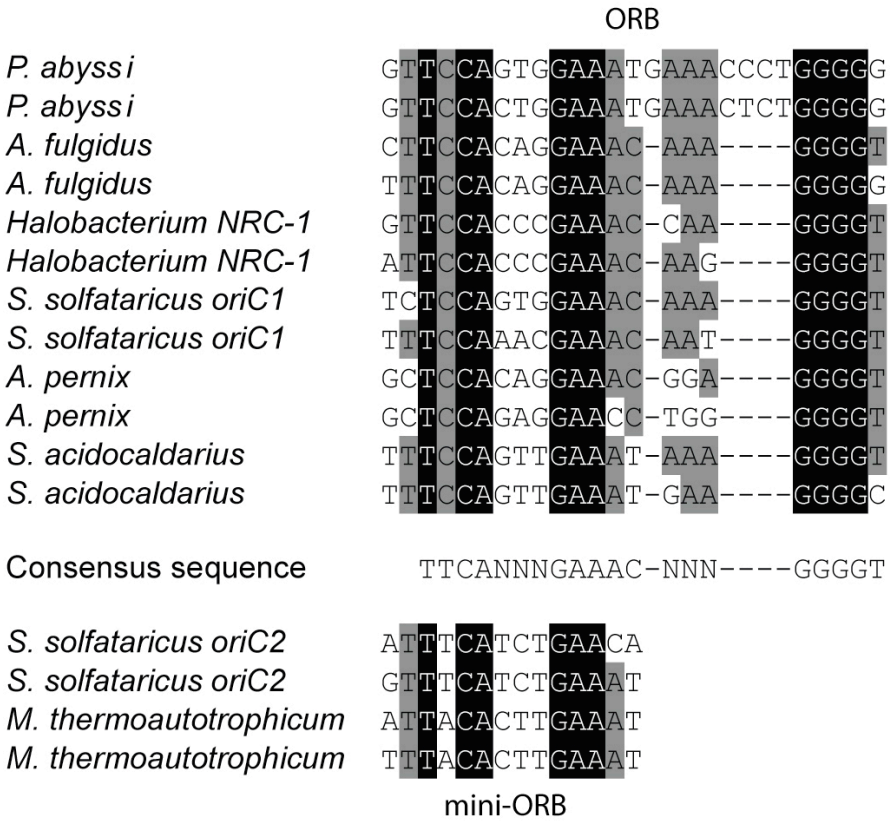


Figure 1.15: ORB and mini-ORB sequences
Alignment of selected ORB repeats from the origins described in Table 1.1. The full length ORB includes the mini-ORB and a guanidine tetranucleotide known as the "G-string". Based on ORB sequences from the references listed in Table 1.1.

1.5.2 Archaeal initiators

A clear *cdc6/orc1* homologue has been found in all archaea with the exception of *M. jannaschii*, *Methanococcus maripaludis* and *Methanococcus kandleri* (Barry and Bell, 2006). Co-localisation of a *cdc6/orc1* homologue and an origin of replication is common but not essential for origin activity (see Table 1.1). Archaeal *cdc6/orc1* homologues have equal similarity with the helicase loader Cdc6 and the C-terminal half of subunit 1 of the origin recognition complex (Barry and Bell, 2006). This similarity suggests that these genes had a common ancestor. *Halobacterium* sp. NRC-1 has multiple *cdc6/orc1* homologues and interestingly not all are essential so they may have overlapping roles or some could be non-functional (Berquist *et al.,* 2007). The first evidence for Cdc6/Orc1-origin interaction came from chromatin

immunoprecipitation studies (ChIP) in *P. abyssi* (Matsunaga *et al.,* 2001). Detailed binding studies of *S. solfataricus* Cdc6/Orc1 proteins have shown that they each bind a subset of *S. solfataricus* origins by interacting with various ORB elements. Different patterns of Cdc6/Orc1-origin binding and Cdc6/Orc1 expression levels may indicate scope for differential origin regulation.

Multiple crystal structures of Cdc6/Orc1 and Cdc6/Orc1 complexed with ATP, ADP or DNA from various archaea have revealed the biochemical details of the initiator-replicator interaction [*Pyrobaculum aerophilum* Cdc6/Orc1, (Liu *et al.,* 2000); *A. pernix* Cdc6/Orc1-ATP/ADP, *A. pernix* Cdc6/Orc1-ORB DNA (Gaudier *et al.,* 2007); *S. solfataricus* Orc1-1 and Orc1-3 bound to *oriC2* DNA (Dueber *et al.,* 2007)]. These structures confirmed the sequence analysis prediction that domains I and II form an AAA+ nucleotide binding fold module at the N-terminus. Domain III contains a functionally crucial winged-helix (WH) fold with significance for DNA binding. Despite being weakly related to DnaA at the amino acid level (Edgell and Doolittle, 1997), the combination of an AAA+ domain and DNA binding domain is similar to DnaA and structural similarity is highly conserved.

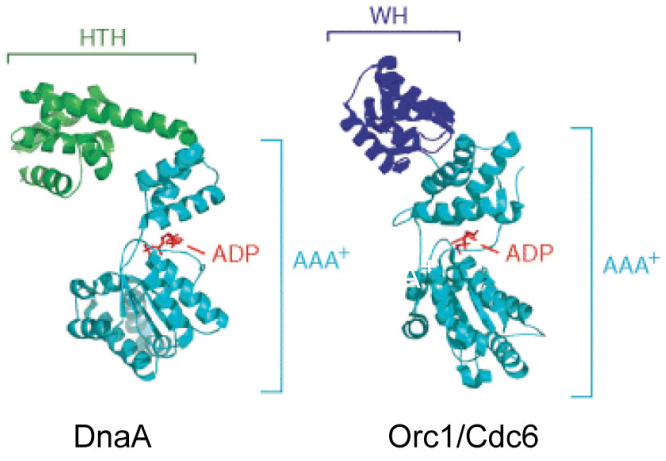


Figure 1.16: Initiator proteins in bacteria and archaea
Structure of bacterial and archaeal initiator proteins showing their structural homology. From Robinson and Bell 2005.

The *A. pernix* ORC-DNA complex exists in a different conformation depending on whether ATP or ADP is bound, this is suggestive of a means of binding regulation. The Cdc6/Orc1-DNA complex structures showed one monomer binding each ORB resulting in local bending of DNA (20-35°) without breaking the base pair hydrogen bonds. WH proteins contain three α-helices and three to four *β*-sheets. The key parts

for DNA binding are two α-helices connected by a loop which make up the helix-turn-helix (HTH) motif, and the "wing" which is a loop connecting the last two *β*-sheets (Tada *et al.,* 2008). The wing motif deeply penetrates a minor groove of the ORB in addition to major groove binding of the α-helix. Phylogenetic analysis of *cdc6/orc1* genes has led to classification based on whether the helix or wing motif respectively shows more pronounced sequence conservation (Berquist and DasSarma, 2003). The published structures show that the two classes may have different DNA binding properties based on the helix/wing-DNA affinity. The loop between the α-helices mediates an interaction between the AAA+ domain and conserved guanine residues at the 3′ end of the ORB sequence. This direct interaction of the AAA+ domain with DNA was unexpected and might be common to all initiators because the α-helix connecting loop is conserved in DnaA and eukaryotic ORC subunits 1, 4, 5 and 6 (Iyer *et al.,* 2004; Tada *et al.,* 2008).

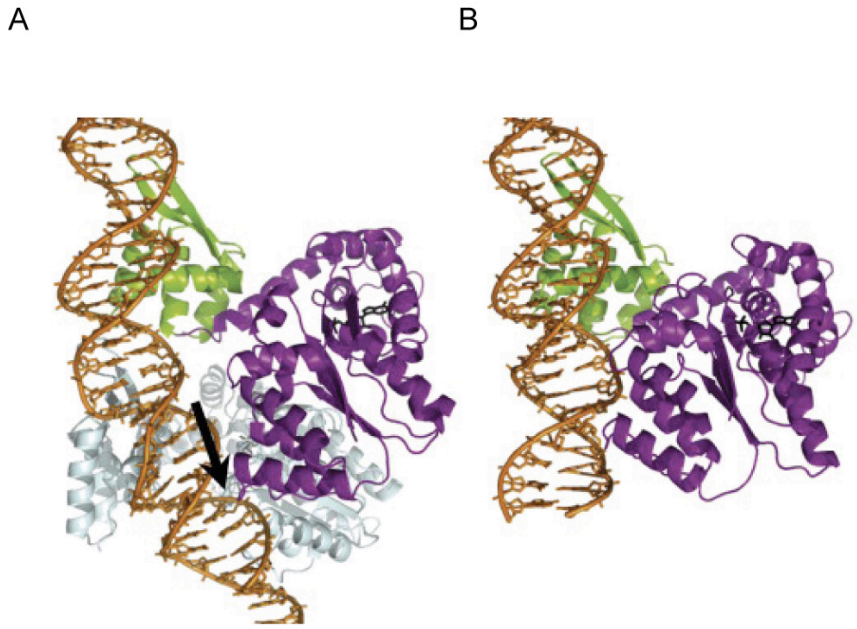


Figure 1.17: Archaeal DNA-Orc1 complex crystal structures
WH domain (green), AAA+ domain (purple). Orc 3 (grey)
(A) Each *A. pernix* Orc1 domain inserts in to minor grooves of the ORB.
(B) The *S. solfataricus* WH domain binds to a minor groove in the ORB sequence and the AAA+ domain inserts in to a major groove. From Wigley, 2009.

At most archaeal origins two inverted ORB elements bookend the DUE. The combination of Cdc6/Orc1-ORB contacts in the winged helix domain and the AAA+ domain give directionality to the interaction and orient Cdc6/Orc1 so that the WH domain faces the DUE. The winged helix domain has been shown to functionally

interact with MCM (see 1.5.3), so orienting MCM towards the DUE explains coordination of the local unwinding of this region during initiation.

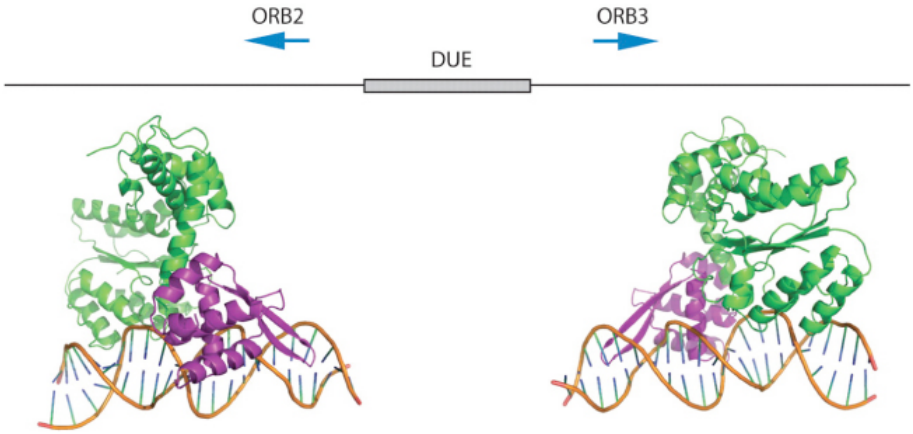


Figure 1.18: *Aeropyrum pernix* Orc bound to ORB elements
Structure of archaeal Orc proteins bound to ORB elements demonstrating the orientation of the WH (purple) domain towards the DUE. From Gaudier *et al* 2007.

The relative paucity of sequence-specific contacts in the initiator-replicator binding suggests that recognition is context-dependent and relies upon replicator structure rather than sequence (Georgescu and O'Donnell, 2007).

Although it is possible that another protein acts as the archaeal initiator, the above evidence points to Cdc6/Orc1 being the likely candidate.

1.5.3 Replisome formation

After the initiator has bound an origin the next step is to recruit the replicative machinery. Little is known about the helicase loading process in archaea. One hypothesis is that the ATPase activity of *cdc6/orc1*, which does not seem to be used for DNA binding, is important for loading (Barry and Bell, 2006). In species with multiple *cdc6/orc1* homologues it is possible that specialisation occurs with some gene products functioning as initiators and others as helicase loaders. Alternatively another protein could perform this function. One candidate is the recently identified WhiP (winged-helix initiator protein). This protein has "modest homology" to *S. cerevisiae* Cdt1 and was found to bind each replication origin of *S. solfataricus* (Robinson and Bell, 2007).

The putative archaeal replicative helicase is an MCM homologue. Befitting of this role, MCM has been shown to localise at archaeal origins *in vivo* (Matsunaga *et al.,* 2001; Matsunaga *et al.,* 2007, Duggin *et al.,* 2008a). Most archaea contain a single MCM which forms a homohexamer *in vitro*. The crystal structure of *M. thermoautotrophicum* MCM shows that the hexamer exists in a ring conformation with a positively charged channel large enough for ds/ssDNA to fit through (Fletcher *et al.,* 2003). Several archaeal MCMs have been characterised biochemically and shown to have 3′-5′ helicase activity in the hexameric ring form (Shin *et al.,* 2009), however the mechanism of MCM DNA unwinding has yet to be described. *P. abyssi* MCM has been shown to bind the *oriC* region in a growth phase-specific manner which is consistent with its putative licensing factor role (Matsunaga *et al.,* 2007). A variety of interactions between MCM and *cdc6/orc1* have been demonstrated *in vitro* in different species. MCM-Cdc6 interactions have been shown to inhibit MCM ATPase activity (De Felice *et al.,* 2003), MCM unwinding activity (Kasiviswanathan *et al.,* 2005) and Cdc6 DNA binding (Kasiviswanathan *et al.,* 2006). Determining the relevant *in vivo* interactions and their consequences is one of the big challenges of archaeal genetics and the next step towards developing a complete model of initiation. Archaea possess homologues of most of the eukaryotic replication proteins and the general mode of replisome formation and fork elongation is thought to be similar to that in eukaryotes (reviewed in Grabowski and Kelman, 2003; Barry and Bell, 2006 and summarised in Figure 1.20).

1.5.4 Origin regulation

Eukaryotes and archaea have multiple opportunities to regulate origin initiation. For example the already mentioned conformation changes associated with initiator-ATP/ADP cycling provide a means of regulating initiator-replicator binding. The *in vitro* modulation of Cdc6/Orc1 autophosphorylation by MCM may play a role in this (Kasiviswanathan *et al.,* 2005). Other possibilities include modulation of chromatin structure (reviewed in Barry and Bell, 2006) and archaeal-specific regulators that might exist within the set of unknown archaeal open reading frames. The possibility of archaeal-specific cell cycle regulators exists because no homologues to the eukaryotic cell cycle regulators CDK or cyclins have been found amongst Archaea (Majernik *et al.,* 2004).

Of particular interest is how the regulation of multiple origins of replication on a single circular chromosome is coordinated. The three origins of replication on the chromosome of *S. acidocaldarius* and *S. solfataricus* initiate synchronously (Lundgren *et al.,* 2004), prompting the authors to suggest that origins might co-localise at a replication factory and/or be co-regulated. More precise characterisation and modelling of *S. acidocaldarius* origin firing showed that *oriC2* replicates slightly later than *oriC1* and *oriC3* (Duggin *et al.,* 2008a). Because these origins are distributed asymmetrically and the replication forks proceed at the same rate termination events occur asynchronously. ChIP showed Cdc6/Orc1 proteins localised at origins throughout the cell cycle in *S. acidocaldarius* while MCM is enriched at origins only during G1 and early S phase (Duggin *et al.,* 2008a). All archaea with multiple *cdc6/orc1* homologues have at least one representative of class I and class II, therefore positive and negative regulation of origins by different Cdc6/Orc1 proteins might be a common mode of regulation. Archaea share a common origin recognition process using eukaryotic-like replication proteins. The structure of archaeal origins resembles the bacterial *oriC* but the existence of multiple origins of replication is similar to eukaryotic chromosomes. Much remains to be done to fully describe the mechanics of replisome formation and origin regulation in the Archaea.

1.6 Elongation

After initiation at a replication origin and assembly of the replisome proteins, the replisome machinery replicates DNA bidirectionally. Ever since the establishment of complementary base pairing in DNA, a "possible copying mechanism for the genetic material" has been conceivable (Watson and Crick, 1953). The mechanism and major proteins that accomplish semi-conservative DNA replication are summarised in Figures 1.19 and 1.20. Although many aspects of DNA replication have been extensively described: uncertainty about the eukaryotic replicative helicase, recent discoveries like the GINS complex and the many unknowns concerning archaeal DNA replication, demonstrate that there are still extensive questions in this field.

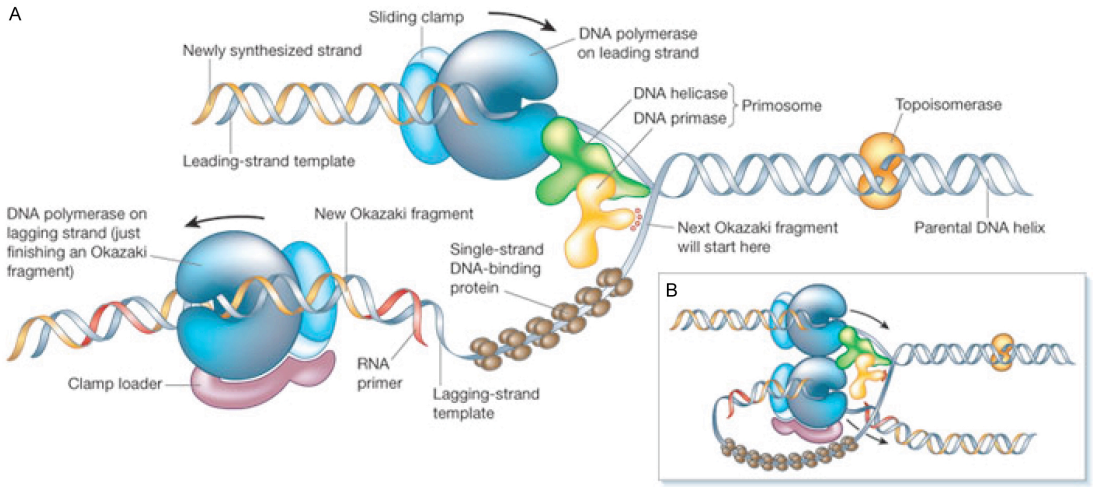


Figure 1.19: DNA replication fork
(A) Replication fork showing the main replisome proteins and the process of DNA replication, see text for a summary of protein functions. Based on the *E. coli* replication fork.
(B) Three-dimensional structure of a replication fork. From Alberts, 2003.

The key functional points for each protein depicted in the replication fork shown in Figure 1.19 are listed below. The principles and functional divisions of DNA synthesis are conserved but in each domain different proteins perform equivalent functions. The different replication factors for each domain are summarised in Figure 1.20, panel D.

Replicative helicase: Unwinds the DNA duplex to expose ssDNA that can be copied by the polymerase.

Single-strand DNA binding protein: Binds and protects ssDNA from degradation while the bases are exposed.

DNA polymerase: Synthesises a complementary DNA strand from parental template by catalysing the polymerisation of deoxyribonucleotides. Polymerases synthesise DNA in the 5′-3′ direction so on the lagging strand short Okazaki fragments are synthesised while on the leading strand continuous synthesis occurs.

Sliding clamp: Anchors DNA polymerase to template DNA to prevent dissociation and vastly increases DNA polymerase processivity.

Clamp loader: Loads and disassembles the sliding clamp.

Primase: Synthesises short RNA primers on the lagging strand from which DNA polymerase can prime.

RNaseH: Removes RNA primers from Okazaki fragments.

DNA ligase: Joins Okazaki fragments on the lagging strand.

Topoisomerase: Regulates DNA supercoiling surrounding the replication fork.

Figure 1.20 summarises the previous sections introducing DNA replication initiation and elongation.

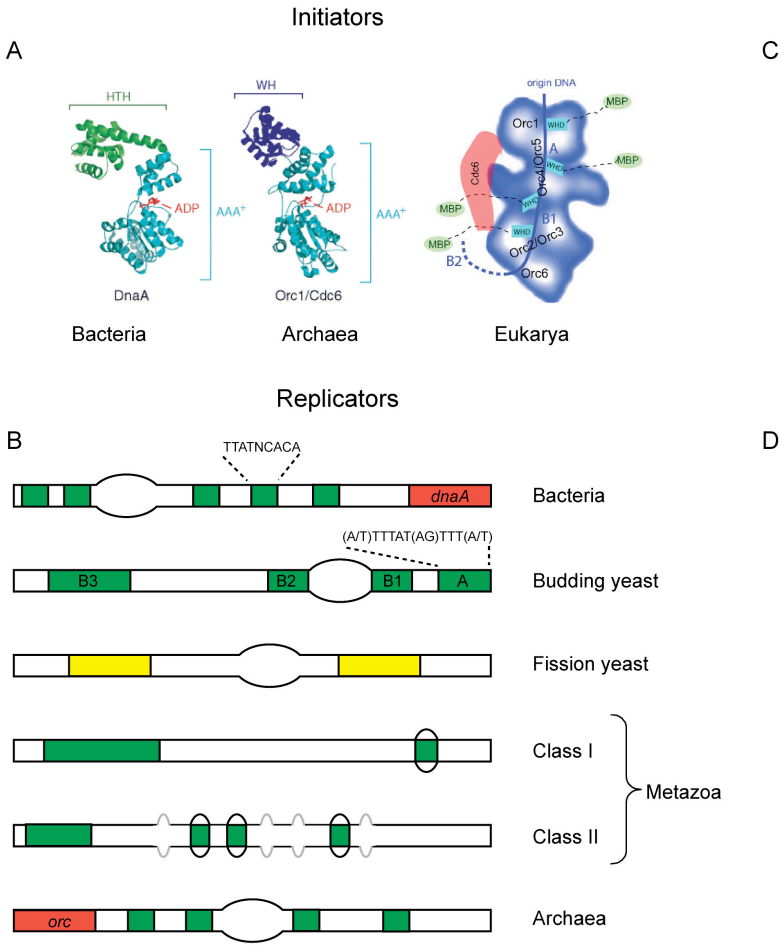
Figure 1.20: Replicators, initiators and replication machinery across the three domains of life

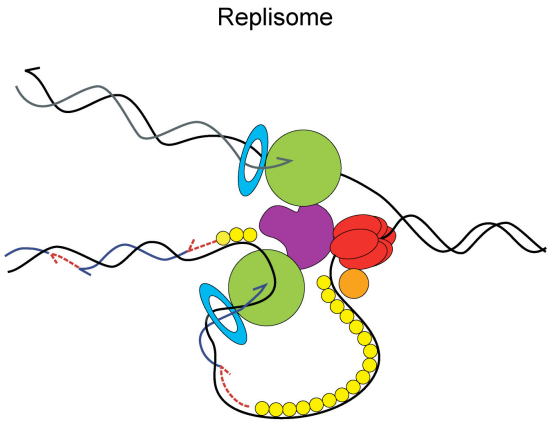
(A) Crystal structures of bacterial and archaeal initiator proteins showing structural homology (from Robinson and Bell, 2005) and a model of the eukaryotic ORC binding to origin DNA (from Chen *et al.,* 2008).

(B) Simplified schematic representation of generalised origins. Bubbles indicate origin sites, green rectangles represent replicators/initiator binding sites (bacteria/budding yeast/archaea) or sequence elements that regulate origin firing or act as sites of preferred origin activity (metazoa). Yellow rectangles represent A/T-rich regions. The replication bubbles are also A/T-rich for archaea, bacteria and budding yeast. Red rectangles represent genes which encode initiator proteins.

(C) Bacterial replisome with colours representing replication proteins described in *E. coli* to give an overview of the protein machinery that accomplishes elongation. Adapted from McGlynn and Lloyd, 2002. Colours refer to panel D.

(D) Replication factors in the three domains of life based on Grabowski and Kelman, 2003 and Barry and Bell, 2006. Many entries are not derived from direct functional evidence. Colours refer to panel C.





D

Function	Bacteria	Eukarya	Archaea	Notes
Origin recognition	DnaA	ORC		Archaeal Orc is similar to subunit 1 of the six subunit eukaryotic ORC and shows equal similarity to cdc6.
Helicase loader	DnaC/DnaI	ORC		DnaI in gram-positive bacteria, DnaC in gram-negative bacteria.
Replicative helicase	DnaB	MCM		Eukaryotes have six MCM subunits, archaea have one.
ssDNA stabilisation	SSB ssDNA binding protein	RPA Replication protein A	RPA/SSB	Euryarchaea have one or three RPA homologues, crenarchaea have a single SSB.
RNA primer synthesis	DnaG	Primeosome	Primase	Archaeal primase has homology to the PriS and PriLeukaryotic subunits.
DNA polymerase	Polymerase III	Polymerase δ and ε	Polymerase B and D	Polδ has three/four subunits, Polε has five subunits. PolB has one and PolD has two subunits. PolB is only in euryarchaea.
Clamp loader	γ-complex	Replication factor C		Eukaryotes have five RFC subunits, archaea have one.
Sliding clamp	β-clamp	Proliferating cell nuclear antigen		Euryarchaea have one subunit and crenarchaea have three.
Fork establishment and progression		GINS complex		Eukaryotes have four GINS proteins and archaea have two, they are a highly divergent family of paralogues. GINS function is yet to be completely characterised.
Primer removal	Polymerase I RNase H	Flap endonuclease I RNase H		
Lagging strand maturation	DNA ligase	DNA ligase I		Bacterial DNA ligase is NAD-dependent, the archaeal and eukaryotic ligase is ATP-dependent
Topoisomerase	Type I and II Reverse gyrase	Type I and II		Reverse gyrase is present in hyperthermophiles and some thermophiles.

Figure 1.20: Replicators, initiators and replication machinery across the three domains of life

1.7 Termination of DNA replication

Once elongation has completed the duplication of an entire replicon, the replication forks collide and the daughter chromosomes are resolved by topoisomerase action. Bacterial termination is only well-understood in *E. coli* or *Bacillus subtilis*. In *E. coli*, termination occurs opposite *oriC* and marks the region where the bidirectional replication forks meet. Inverted repeats called *Ter* sequences function in a polar manner and block replication forks travelling from a specific direction when bound by the Tus terminator protein (Kornberg, 1992). This is achieved by a tight Tus-*Ter* complex which can not be dislodged by an approaching replicative helicase from a specific direction (Mulcair *et al.,* 2006). Opposed groups of *Ter* sites form a replication trap that ensures replication forks can only enter the terminus region (reviewed in Duggin *et al.,* 2008b). Termination in *B. subtilis* operates using the same principle but neither the *Ter* sites or the terminator protein show sequence or structural homology.

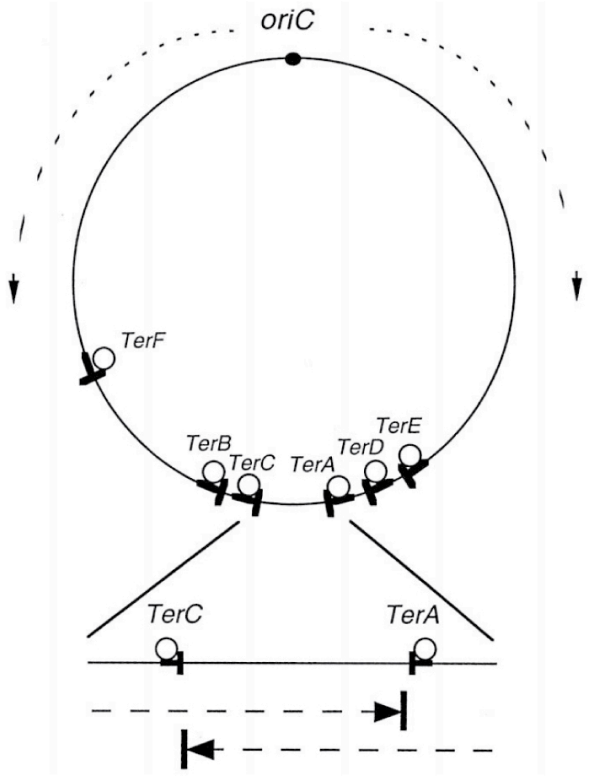


Figure 1.21: Terminus region of *Escherichia coli*
Replication forks originating at *oriC* are shown by dashed lines. The circles at *Ter* sites represent Tus monomers and the T shape of the *Ter* sites denotes their polarity. The enlargement of the termination region shows how opposed *Ter* sites act as a replication trap. From Hyrien, 2000.

In eukaryotes the confusion about replicators has limited study of the termination process. The exception is the end-replication of linear chromosome ends using the well characterised telomerase synthesis of telomeres (reviewed in Gilson and Geli, 2007). Multiple origins of replication in eukaryotes mean that the majority of replication bubbles merge upstream and downstream with forks from neighbouring origins. Given the flexibility of origin firing and timing it is unlikely that sequence-specific termination occurs because forks will not meet at conserved locations. However there are exceptions where termination can be studied such as the replication fork barrier associated with rDNA. The best characterised example is in *S. cerevisiae* where the polar terminus is at the 3′ end of the rDNA to ensures that rDNA replication is unidirectional (Brewer and Fangman, 1988). With the study of archaeal DNA replication barely a decade old, few inroads have been made in to understanding termination in the third domain of life.

1.8 Origin-independent replication

DNA replication, repair and recombination are sometimes called “the three Rs” to highlight their importance. All three processes are crucial in maintaining genome stability and the extent of their interdependence is emphasised in current research. For example, in nucleotide excision repair the gaps produced by excising damaged DNA need to be filled by repair synthesis action of a DNA polymerase. This chapter has so far dealt with an idealised version of DNA replication where it proceeds unhindered directly from origin to terminus. However, this is not always the case and replication forks will often encounter blockages that cause fork stalling or collapse. Examples of blockages include single-stranded nicks in the phosphodiester backbone, double-strand breaks and bound proteins such as nucleosomes or stalled RNA polymerases (McGlynn and Lloyd, 2002). One low fidelity method of bypassing blockages is by extension across the region of nucleotide damage using translesion polymerases. This involves switching to a translesion polymerase and then returning to the standard replicative polymerase for the resumption of normal DNA synthesis (Friedberg *et al.,* 2005). Translesion polymerases are found in all domains of life but they have high nucleotide misincorporation rates. The high mutation rates generated by translesion polymerase synthesis are avoided by using the recombination machinery to repair or bypass blockages (See Figure 1.22) (Berdichevsky *et al.,* 2002).

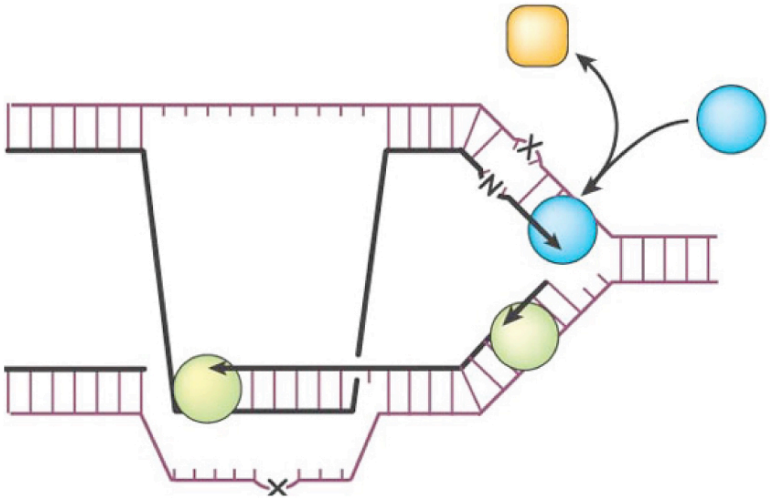


Figure 1.22: Replication fork block
Replication fork blockages (X) can be bypassed using a translesion polymerase (orange) or by template switching and use of a homologous template to replicate damaged DNA (bottom strand). From Hoeijmakers, 2001.

1.8.1 Recombination

Recombination describes the process where DNA molecules interact resulting in the rearrangement of genetic information. It generates genetic diversity and has an important role in DNA repair and replication. Recombination can be divided in to homologous recombination, DNA transposition and illegitimate recombination.

Homologous recombination is the genetic exchange between strands of DNA that share regions of sequence identity. Recombination can be used to repair deleterious lesions such as double-strand DNA breaks. Double-strand breaks are resected by exonucleases which exposes 3′-ended ssDNA tails. In homologous recombination the ssDNA substrate is bound by the recombinase protein to form an extended nucleoprotein helical filament. The recombinase then catalyses a homology search and strand invasion. The mechanism by which homology is sought is unknown. The presynaptic filament binds duplex DNA forming a triple-stranded D-loop structure from which replication can prime (Figure 1.23 and reviewed in Sung and Klein, 2006).

In all domains of life, recombinases are ssDNA-dependent ATPases that bind DNA with no sequence specificity. They are essential for catalysing strand exchange. The bacterial recombinase is RecA. Rad51 is the general eukaryotic strand exchange protein and Dmc1 is a eukaryotic recombinase specific to meiotic recombination.

The archaeal recombinase shares 40% identity with Rad51 and 20% with RecA and is called RadA (Sandler *et al.,* 1996). These recombinases share a core domain containing conserved ATP binding and hydrolysis motifs but are dissimilar at their extremes. RadA, Dmc1 and Rad51 share a conserved N-terminus while RecA has an extra C-terminal region (Komori *et al.,* 2000). Homologous recombination is used in a multitude of repair pathways.

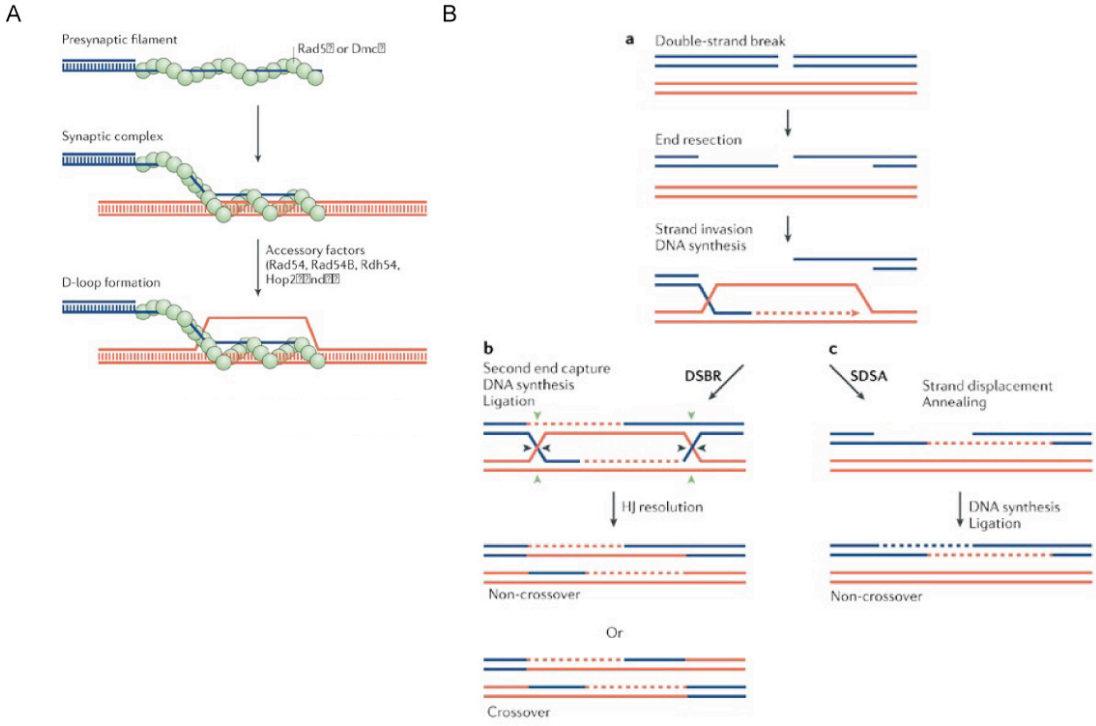


Figure 1.23: Homologous recombination
(A) Eukaryotic model of recombinase assembly on ssDNA and subsequent D-loop formation through strand invasion. From Sung and Klein, 2006.
(B) The double-strand break repair (DSBR) and synthesis-dependent strand annealing (SDSA) models for repairing double-strand breaks from a homologous recombination D-loop substrate. From Sung and Klein, 2006.

Site-specific recombination is an example of DNA transposition which involves no DNA synthesis and relies upon a specialised recombinase that recognises particular sequences. The most well known example is lambda bacteriophage which uses an *attP* sequence to recombine with bacterial *attB* sites. Intermolecular site-specific recombination can result in integration of circular molecules while intramolecular site-specific recombination leads to either sequence excision or inversion depending upon the orientation of the recognition sequences (Grindley *et al.,* 2006). Mobile genetic elements are also examples of DNA transposition. Illegitimate recombination

does not fit in to any of the above categories. One example is known as non-homologous end-joining and can be used to repair double-strand breaks. Since this process is not based on homology it can occur between random double-strand breaks and is potentially mutagenic.

1.8.2 Replication restart

Replication from sites other than *oriC* has been observed in *E. coli* and is divided in to induced-stable DNA replication (iSDR) and constitutive-stable DNA replication (cSDR). These types of aberrant replication are distinguishable by their protein requirements and origin sites (summarised in Table 1.2).

Table 1.2: Comparison of *Escherichia coli* replication systems
Replication system requirements and properties. Adapted from Kogoma, 1997.

Replication system	Cell type	Duplex opening mode	Origin	DnaA	RecA	RecBC	PriA	Protein synthesis	RNA synthesis
DnaA/*oriC*	Normal	DnaA-*ori*C interaction	*oriC*	+	-	-	-	+	+
iSDR	SOS induced	D-loop	*oriM* sites	-	+	+	+	-	-
cSDR	*rnhA recG* mutants	R-loop	*oriK* sites	-	+	-	+	-	+

The dependence of iSDR and cSDR on recombination proteins such as RecA (Masai *et al.,* 1994), and their independence of *oriC*, led Kogoma and colleagues to suggest that recombination mechanisms could be used to initiate replication at stalled replication forks. Since stalled replication forks can occur at any sequence, replication restart proteins have to be capable of recognising structural elements of forks. PriA has been found to act upon D-loops formed as a result of homologous recombination at stalled replication forks. PriA enables replication restart by beginning a cascade of protein interactions that result in DnaB recruitment (reviewed in McGlynn and Lloyd, 2002). The amount of replisome dissociation at collapsed forks is unknown but DnaB reloading is required for replication restart so PriA is a key coordinator of replication restart (Rudolph *et al.,* 2007). It is difficult to estimate the frequency of replication fork collapse but cell death in strains deficient for recombination implies that one of the main functions of recombination is to restart stalled or collapsed replication forks. The restart of replication at recombination

intermediates resulting from DNA repair uses an origin-independent method of initiating DNA replication called recombination-dependent replication (RDR). Other models for replication-restart are shown in Figure 1.24.

Less is known about replication restart in eukaryotes and archaea. In contrast to bacteria, it is thought that eukaryotic replication forks are actively stabilised to prevent dissociation of the replisome machinery (Lopes *et al.,* 2001). There seems to be a variety of pathways that can be used to initiate stalled forks but the molecular details are not yet known and probably depend upon the nature of the blockage (Rudolph *et al*. 2005; Lambert *et al.,* 2007). In eukaryotes multiple replication origins help to ensure complete replication because regions of DNA not replicated by stalled forks can be replicated by forks emanating from adjacent origins (McGlynn and Lloyd, 2002). In eukaryotes cell-cycle checkpoints are triggered by stalled replication forks (Lambert and Carr, 2005). Cell-cycle checkpoints activate DNA repair pathways, preventing late origins from firing, delaying the cell-cycle and by initiating replisome stabilisation mechanisms (reviewed in Branzei and Foiani, 2007). Few homologues of bacterial replication restart have been identified in eukaryotes. Break-induced replication involves template switching and may be important for replication restart in eukaryotes (Llorente *et al.,* 2008). Homologues of helicases involved in replication restart are mostly absent from archaea so they may have novel proteins for processing stalled replication forks .

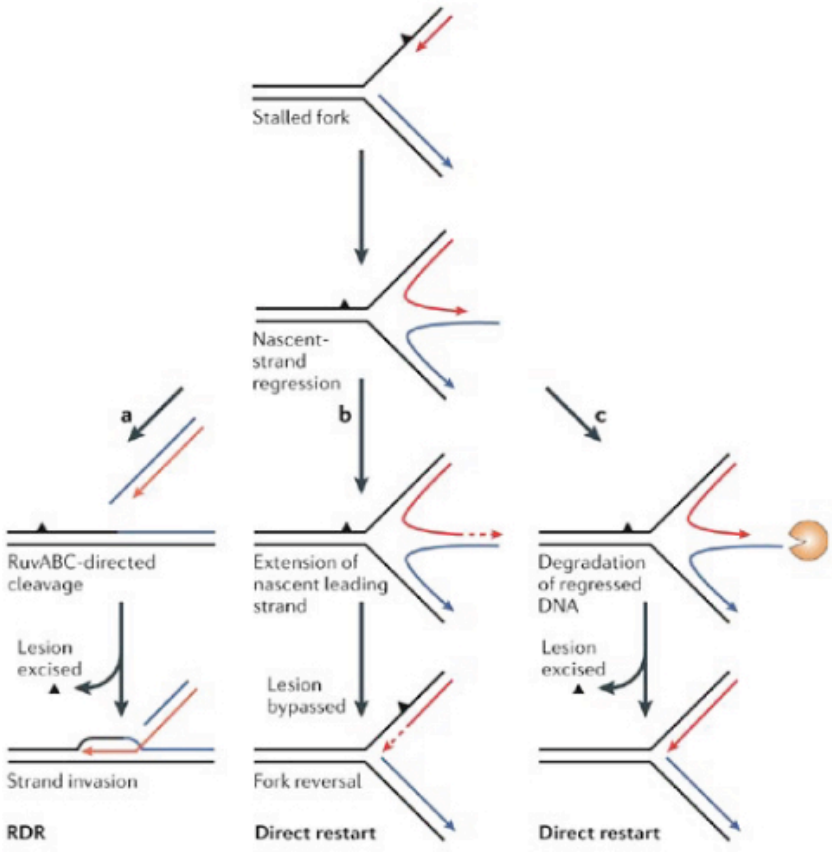


Figure 1.24: Replication restart mechanisms in bacteria

When a replisome encounters a leading-strand template lesion (black triangle) the stalled fork undergoes nascent-strand regression to form a Holliday-junction, replication can then be restarted by:

(A) RDR pathway where the Holliday junction is resolved by the branch-migration helicase and Holliday-junction endonuclease RuvABC. After excision of the template lesion, the broken chromosome undergoes recombination with the intact duplex to form a D-loop structure and replication restart takes place.

(B) In the template-switching model, strand exchange (Figure 1.22) or fork regression and annealing are used. The nascent leading strand is extended, using the nascent lagging strand as a template. After reversal to a fork structure, direct restart occurs.

(C) Nascent-strand regression allows for excision of the template lesion. The regressed DNA of both the nascent leading and lagging strand is degraded by an exonuclease and direct replication restart takes place.

Black strands, parental DNA; red strands, nascent leading-strand DNA; red dotted strands, extension past the blocking lesion; blue strands, nascent lagging-strand DNA. The arrowheads at the end of the nascent DNA represent the 3′-OH termini. From Heller and Marians, 2006.

1.9 *Haloferax volcanii* as a model organism

Haloarchaea require a hypersaline environment and maintain a high intracellular concentration of inorganic salts (Soppa, 2006). *H. volcanii* (formerly known as *Halobacterium volcanii*) is a halophilic euryarchaeote initially isolated from the Dead Sea (Mullakhanbhai and Larsen, 1975).

Table 1.3: *Haloferax volcanii* properties
Characterisation of *H. volcanii* from Mullakhanbhai and Larsen, 1975

Property	Description
Morphology	Disc shaped cells 1-3×2-3 μm, 0.4-0.5 μm in thickness. Non-motile and pleomorphic.
Culture	Chemoorganotrophic, aerobic. No growth anaerobically with nitrate
Chemical Composition	Large excess of acidic over basic amino acids in bulk protein. Amino sugar content of cell envelope very low; cell envelope lacks muramic acid. Lipid largely terpenoid; no fatty acid derivatives present. Pigmentation red-to-orange due to carotenoids.

H. volcanii grows aerobically in media containing 2.5 M NaCl at 45°C (optimum) and has a wide tolerance of growth conditions, growing between 30°-50°C and from 1 M NaCl to saturation. The generation time has been reported as 3-4 hours (Halohandbook [www.haloarchaea.com/resources/halohandbook]).

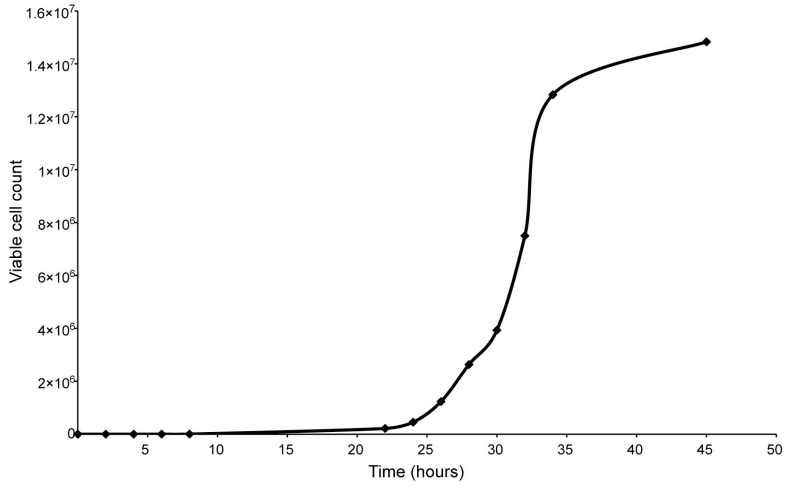


Figure 1.25: *Haloferax volcanii* growth curve
Example viable cell count for *H. volcanii* strain H53 (Δ*pyrE2*, Δ*trpA*) grown at 45°C in YPC media containing 2.5M NaCl. Calculated generation time from an average of three growth curves is 2.8 hours.

H. volcanii has a multi-replicon genome of ~4 Mb and a G/C content of ~66%. Its genome consists of a main circular chromosome (2.9 Mb) and four smaller replicons: pHV1 (86 kb), pHV3 (442 kb), pHV4 (690 kb) and pHV2 (6.4 kb) (Charlebois *et al.,* 1991). *H. volcanii* is polyploid and the genome copy number varies from 10 copies per cell during stationary phase to 18 copies per cell during exponential phase (Breuert *et al.,* 2006).

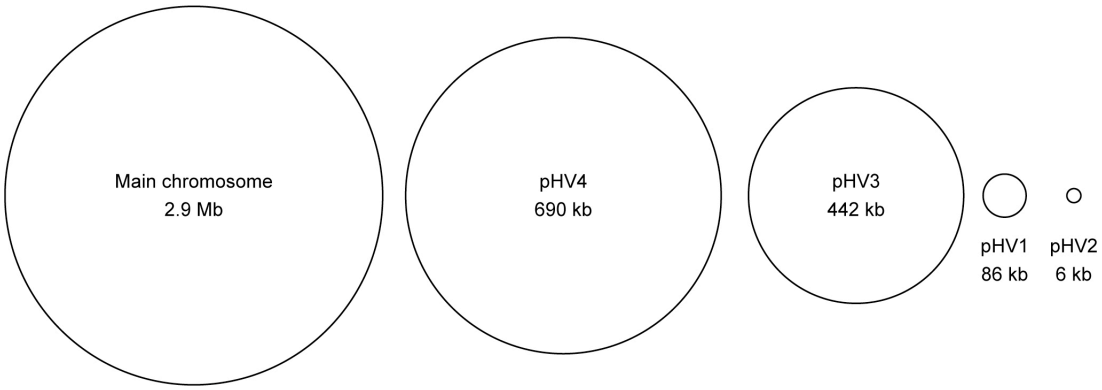


Figure 1.26: *Haloferax volcanii* replicons
Main chromosome and pHV2 are not to scale. pHV2 is not present in DS70 or WFD11 laboratory strains.

The wild-type *H. volcanii* strain is known as DS2. pHV2 contains six open reading frames (ORF) encoding hypothetical and unknown proteins. The largest ORF encodes a homologue of a gene called *repH* that is necessary for replication of a 200 kb plasmid of *Halobacterium halobium* (Ng and DasSarma, 1993). Therefore pHV2 presumably replicates using a RepH-dependent replication origin. Both of the two laboratory strain backgrounds of *H. volcanii* derived from DS2 have been cured of pHV2. WFD11 background pHV2 curing was accomplished using ethidium bromide treatment (Charlebois *et al.,* 1987). Due to this mutagenic process, WFD11 strains contain many rearrangements and are not ideal for genetic studies. The DS70 strain background was generated using plasmid incompatibility to cure pHV2 (Wendoloski *et al.,* 2001). These laboratory strains enabled the development of transformation protocols and pHV2-based shuttle vectors (Charlebois *et al.,* 1987).

Haloarchaea have generally proved more amenable to genetic analysis than most archaea. The characteristics of *H. volcanii* mean that it can be easily cultured in the laboratory on agar-based media. Adaptation of standard genetic techniques has enabled *H. volcanii* to become a well-established model system for archaea (Soppa,

2006). One of the foundations of genetic analysis is the transformation of cells. This is possible in *H. volcanii* using PEG-mediated transfer (Cline *et al.,* 1989) and there are a variety of antibiotic and auxotrophic selectable markers available (Allers *et al.,* 2004; Allers and Mevarech, 2005). The constitutive *fdx* promoter ensures expression of these markers in transformants (Gregor and Pfeifer, 2005). Inducible promoters have recently been developed and have already proved useful in the analysis of an essential gene (Large *et al.,* 2007). The large size of the pHV2 origin on the original shuttle vectors and the inability of these vectors to function in a Δ*radA* strain (Woods and Dyall-Smith, 1997), meant that shuttle plasmid vectors based on other *H. volcanii* origins have also been of value (Norais *et al.,* 2007). These protocols are all components of a method developed to construct deletion strains for genetic analysis (Bitan-Banin *et al.,* 2003). This method is described in full in section 2.2.6.

The *H. volcanii* genome has been sequenced by The Institute for Genomic Research (TIGR) and is available through the UCSC Archaeal Genome Browser (http://archaea.ucsc.edu/), with the annotation to follow (Hartman *et al*., in preparation). Several haloarchaeal genomes have now been sequenced. Common features are multiple copies of genes that are single-copy in other archaea and multiple insertion elements (Soppa, 2006).

The genetic tools, ease of culture and the genome sequence resource make *H. volcanii* an excellent model organism for studying DNA replication.

1.10 Objectives

The main focus of this work was to characterise the genome replication system of *H. volcanii*. This involved:

- Identifying DNA replication origins for each replicon.
- Describing replication origin properties and features.
- Investigating the function of a gene family found near a replication origin locus.

A study of replication origins in *H. volcanii* increases understanding of DNA replication within the Archaeal domain of life, specifically the Euryarchaeota phylum. Identifying DNA replication origins in *H. volcanii* should help begin addressing the question of multi-replicon genome evolution. In addition it has the potential to generate useful tools for the genetic manipulation of *H. volcanii*.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Tables

Table 2.1: *Escherichia coli* strains

Strain	Name	Genotype	Relevant properties
XL1-Blue MRF′	E2	endA1 gyrA96 (NalR) lac [F′ proAB lacIqZΔM15 Tn10 (TetR)] Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 recA1 relA1 supE44 thi-1	Standard cloning vector for blue/white selection using pBluescript based plasmids. Tetracycline used to select against loss of F′ episome
N2338 (GM121)	E3	*F– ara-14 dam-3 dcm-6 fhuA31 galK2 galT22 hsdR3 lacY1 leu-6 thi-1 thr-1 tsx-78*	*dam*$^-$ *dcm*$^-$ mutant for preparing DNA for *H. volcanii* transformations.

Table 2.2: *Haloferax volcanii* strains

For explanation of plasmid integration and pop-out strains see 2.2.6 p77-79

Strain/alias	Parent Strain	Genotype	Notes
H1/WR340	H8	Δ*his*	(Bitan-Banin *et al.,* 2003)
H9/DS70	H11	ΔpHV2	(Wendoloski *et al.,* 2001)
H11/H678 /DS2	n/a	WT	(Mullakhanbhai and Larsen, 1975)
H18	H9	*pyrE2*+::[Δ*pyrE2* Nov^R]	(Allers *et al.,* 2004)
H26	H18	Δ*pyrE2*	(Allers *et al.,* 2004)
H49	H29	Δ*his*, Δ*pyrE2*, Δ*radA*	(Norais *et al.,* 2007)
H50	H26	Δ*pyrE2*, *bgaHv*::[*bgaH pyrE2*$^+$]	(Delmas *et al*., submitted)
H53	H47	Δ*pyrE2*, Δ*trpA*	(Allers *et al.,* 2004)
H54	H50	Δ*pyrE2*, *bgaHa*	(Delmas *et al*., submitted)
H80	H26	Δ*pyrE2*, *radA*+::[Δ*radA pyrE2*$^+$]	Integration of pTA80 (Δ*radA pyrE2*$^+$) at *radA*. Constructed by Thorsten Allers.
H98	H90	Δ*pyrE2*, Δ*hdrB*	(Allers *et al.,* 2004)
H112	H80	Δ*pyrE2*, Δ*radA*	(Norais *et al.,* 2007)

H535	H26	Δ*pyrE2*, *bgaHv*::[Δ*bgaHa pyrE2*$^+$]	(Delmas *et al*., submitted)
H555	H535	Δ*pyrE2*, Δ*bgaHv*	(Delmas *et al*., submitted)
H884	H112	Δ*pyrE2*, Δ*radA*, *ori-pHV1/4*+::[*ori-pHV1/4 pyrE2*$^+$]	Integration of pTA250 (*ori-pHV1/4 pyrE2*$^+$) at *ori-pHV1/4*.
Origin deletion strains			
H220	H53	Δ*pyrE2*, Δ*trpA*, *ori-pHV1/4*+::[Δ*ori-pHV1/4*::*trpA*$^+$ *pyrE2*$^+$]	(Norais *et al.,* 2007)
H230	H220	Δ*pyrE2*, Δ*trpA*, Δ*ori-pHV1/4*::*trpA*$^+$	(Norais *et al.,* 2007)
H299	H230	Δ*pyrE2*, Δ*trpA*, Δ*ori-pHV1/4*::*trpA*$^+$, *hdrB*+::[Δ*hdrB pyrE2*$^+$]	Integration of pTA160 (Δ*hdrB*, *pyrE2*$^+$) at *hdrB*.
H300	H299	Δ*pyrE2*, Δ*trpA*, Δ*ori-pHV1/4*::*trpA*$^+$, Δ*hdrB*	Pop-out of pTA155 (Δ*hdrB*, *pyrE2*$^+$) from *hdrB* locus, leaving behind Δ*hdrB*. Constructed by Thorsten Allers.
H446	H53	Δ*pyrE2*, Δ*trpA*, *oriC-2*+::[Δ*oriC-2*::*trpA*$^+$ *pyrE2*$^+$]	Integration of pTA531 (Δ*oriC-2*::*trpA*$^+$, *pyrE2*$^+$) at *oriC-2*. Constructed by Thorsten Allers.
H447	H53	Δ*pyrE2*, Δ*trpA*, *oriC-1*+::[Δ*oriC-1*::*trpA*$^+$ *pyrE2*$^+$]	Integration of pTA532 (Δ*oriC-1*::*trpA*$^+$, *pyrE2*$^+$) at *oriC-1*. Constructed by Thorsten Allers.
H471 H472	H446	Δ*pyrE2*, Δ*trpA*, Δ*oriC-2*::*trpA*$^+$	Pop-out of pTA531 (Δ*oriC-2*::*trpA*$^+$, *pyrE2*$^+$) from *oriC-2* locus to give Δ*oriC-2*::*trpA*$^+$. Constructed by Thorsten Allers.
H473 H474	H447	Δ*pyrE2*, Δ*trpA*, Δ*oriC-1*::*trpA*$^+$	Pop-out of pTA532 (Δ*oriC-1*::*trpA*$^+$, *pyrE2*$^+$) from *oriC-1* locus to give Δ*oriC-1*::*trpA*$^+$. Constructed by Thorsten Allers.
H955	H300	Δ*pyrE2*, Δ*trpA*, Δ*ori-pHV1/4*::*trpA*$^+$, Δ*hdrB*, Δ*oriC-1*::[Δ*oriC-1*::*hdrB pyrE2*$^+$]	Integration of pTA946 (Δ*oriC-1*::*hdrB*, *pyrE2*$^+$) at *oriC-1*.

H1023	H955	Δ*pyrE2*, Δ*trpA*, Δ*ori-pHV1/4*::*trpA*$^+$, Δ*hdrB*, Δ*oriC-1*::*hdrB*$^+$	Pop-out of pTA946 (Δ*oriC-1*::*hdrB*, *pyrE2*$^+$) to give Δ*oriC-1*::*hdrB*$^+$.
rad25 deletion strains			
H666	H53	Δ*pyrE2*, Δ*trpA*, *rad25D*::[Δ*rad25D pyrE2*$^+$]	Integration of pTA453 (Δ*rad25D*, *pyrE2*$^+$) at *rad25D*.
H668	H53	Δ*pyrE2*, Δ*trpA*, *rad25B*::[Δ*rad25B pyrE2*$^+$]	Integration of pTA737 (Δ*rad25B*, *pyrE2*$^+$) at *rad25B*.
H669	H53	Δ*pyrE2*, Δ*trpA*, *rad25C*::[Δ*rad25C pyrE2*$^+$]	Integration of pTA738 (Δ*rad25C*, *pyrE2*$^+$) at *rad25C*.
H679	H666	Δ*pyrE2*, Δ*trpA*, Δ*rad25D*	Pop-out of pTA453 (Δ*rad25D*, *pyrE2*$^+$) to give Δ*rad25D*.
H680	H669	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*	Pop-out of pTA738 (Δ*rad25C*, *pyrE2*$^+$) to give Δ*rad25C*.
H709	H668	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*	Pop-out of pTA737 (Δ*rad25B*, *pyrE2*$^+$) to give Δ*rad25B*.
H711	H679	Δ*pyrE2*, Δ*trpA*, Δ*rad25D*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	Integration of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) at *radA*.
H712	H680	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	
H714	H709	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	Integration of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) at *radA*.
H750	H680	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, *rad25D*::[Δ*rad25D pyrE2*$^+$]	Integration of pTA453 (Δ*rad25D*, *pyrE2*$^+$) at *rad25D*.
H752	H709	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, *rad25D*::[Δ*rad25D pyrE2*$^+$]	
H754	H709	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, *rad25C*::[Δ*rad25C pyrE2*$^+$]	Integration of pTA738 (Δ*rad25C*, *pyrE2*$^+$) at *rad25C*.
H773	H754	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25C*	Pop-out of pTA738 (Δ*rad25C*, *pyrE2*$^+$) to give Δ*rad25C*.
H774	H752	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25D*	Pop-out of pTA453 (Δ*rad25D*, *pyrE2*$^+$) to give Δ*rad25D*.

H781	H750	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, Δ*rad25D*	
H783	H774	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25D*, *rad25C*::[Δ*rad25C* *pyrE2*$^+$]	Integration of pTA738 (Δ*rad25C*, *pyrE2*$^+$) at *rad25C*.
H789	H783	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*	Pop-out of pTA738 (Δ*rad25C*, *pyrE2*$^+$) to give Δ*rad25C*.
H791	H773	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25C*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	
H792	H774	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25D*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	Integration of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) at *radA*.
H794	H781	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, Δ*rad25D*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	
H795	H789	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	
H819	H712	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, Δ*radA*::*trpA*$^+$	
H821	H791	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25C*, Δ*radA*::*trpA*$^+$	
H822	H792	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25D*, Δ*radA*::*trpA*$^+$	
H830	H714	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*radA*::*trpA*$^+$	Pop-out of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) to give Δ*radA*::*trpA*$^+$.
H831	H711	Δ*pyrE2*, Δ*trpA*, Δ*rad25D*, Δ*radA*::*trpA*$^+$	
H833	H794	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, Δ*rad25D*, Δ*radA*::*trpA*$^+$	
H834	H795	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*, Δ*radA*::*trpA*$^+$	
H837	H53	Δ*pyrE2*, Δ*trpA*, *rad25A*::[Δ*rad25A pyrE2*$^+$]	
H839	H679	Δ*pyrE2*, Δ*trpA*, Δ*rad25D*, *rad25A*::[Δ*rad25A pyrE2*$^+$]	Integration of pTA882

H841	H680	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, *rad25A*::[Δ*rad25A pyrE2*$^+$]	(Δ*rad25A*, *pyrE2*$^+$) at *rad25A*.
H843	H709	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, *rad25A*::[Δ*rad25A pyrE2*$^+$]	Integration of pTA882 (Δ*rad25A*, *pyrE2*$^+$) at *rad25A*.
H844	H773	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25C*, *rad25A*::[Δ*rad25A pyrE2*$^+$]	
H845	H774	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25D*, *rad25A*::[Δ*rad25A pyrE2*$^+$]	
H846	H781	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, Δ*rad25D*, *rad25A*::[Δ*rad25A pyrE2*$^+$]	
H847	H789	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*, *rad25A*::[Δ*rad25A pyrE2*$^+$]	
H851	H837	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*	Pop-out of pTA882 (Δ*rad25A*, *pyrE2*$^+$) to give Δ*rad25A*.
H852	H839	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25D*	
H853	H841	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25C*	
H854	H843	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*	
H855	H844	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25C*	
H856	H845	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25D*	
H857	H846	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25C*, Δ*rad25D*	
H858	H847	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*	
H872	H851	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	Integration of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) at *radA*.
H873	H852	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25D*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	
H876	H854	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	

H877	H855	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25C*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	Integration of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) at *radA*.
H878	H856	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25D*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	
H879	H857	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25C*, Δ*rad25D*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	
H881	H858	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$]	
H908	H873	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25D*, Δ*radA*::*trpA*$^+$	Pop-out of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) to give Δ*radA*::*trpA*$^+$.
H909	H876	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*radA*::*trpA*$^+$	
H910	H881	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*, Δ*radA*::*trpA*$^+$	
H934	H879	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25C*, Δ*rad25D*, Δ*radA*::*trpA*$^+$	
H990	H851	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*radA*::*trpA*$^+$	
H991	H877	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25C*, Δ*radA*::*trpA*$^+$	
H992	H878	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25D*, Δ*radA*::*trpA*$^+$	
bga mutant strains			
H50	H26	Δ*pyrE2*, *bgaHv*::[*bgaH pyrE2*$^+$]	Integration of pTA102 (*bgaH* from *H. alicantei*, *pyrE2*$^+$) at *bgaHv*. Constructed by Thorsten Allers.
H54	H50	Δ*pyrE2*, *bgaHa*	Pop-out of pTA102 (*bgaH* from *H. alicantei*, *pyrE2*$^+$) from *bgaHv* locus, leaving behind *bgaHa*. Constructed by Thorsten Allers.

H292	H251	Δ*pyrE2*, Δ*trpA*, *bgaHa-Kp*	Pop-out of pTA95 (Δ*trpA*, *pyrE2*$^+$) from *trpA* locus, to give Δ*trpA*. Constructed by Thorsten Allers (H251 described in Delmas et al., submitted)
H320	H292	Δ*pyrE2*, Δ*trpA*, *bgaHa-Kp*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$*]*	Integration of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) at *radA*.
H555	H535	Δ*pyrE2*, Δ*bgaHv*	Pop-out of pTA617 (Δ*bgaHa*, *pyrE2*$^+$) from *bgaHv* locus, to give Δ*bgaHv*. Constructed by Thorsten Allers.
H581	H320	Δ*pyrE2*, Δ*trpA*, *bgaHa-Kp*, *radA*+::[Δ*radA*::*trpA*$^+$ *pyrE2*$^+$*]*, {*radA*$^+$ *pyrE2*$^+$ Mev^R}	H320 with episomal pTA636 (*pyrE2*$^+$ Mev^R *radA*$^+$)
H607	H581	Δ*pyrE2*, Δ*trpA*, *bgaHa-Kp*, Δ*radA*::*trpA*$^+$	Pop-out of pTA324 (Δ*radA*::*trpA*$^+$, *pyrE2*$^+$) to give Δ*radA*::*trpA*$^+$. Constructed by Thorsten Allers.
H994	H679	Δ*pyrE2*, Δ*trpA*, Δ*rad25D*, *bgaHv*::[*bgaH pyrE2*$^+$]	Integration of pTA102 (*bgaH* from *H. alicantei*, *pyrE2*$^+$) at *bgaHv* locus of *H. volcanii*.
H995	H680	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, *bgaHv*::[*bgaH pyrE2*$^+$]	
H996	H709	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, *bgaHv*::[*bgaH pyrE2*$^+$]	
H997	H851	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, *bgaHv*::[*bgaH pyrE2*$^+$]	
H998	H858	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*, *bgaHv*::[*bgaH pyrE2*$^+$]	
H999	H53	Δ*pyrE2*, Δ*trpA*, *bgaHv*::[*bgaH pyrE2*$^+$]	
H1005	H994	Δ*pyrE2*, Δ*trpA*, Δ*rad25D*, *bgaHa*	Pop-out of pTA104 (pTA102 (*bgaH pyrE2*$^+$) from *bgaHv* locus to give *bgaHa*.
H1006	H995	Δ*pyrE2*, Δ*trpA*, Δ*rad25C*, *bgaHa*	
H1007	H996	Δ*pyrE2*, Δ*trpA*, Δ*rad25B*, *bgaHa*	
H1008	H997	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, *bgaHa*	

H1009	H998	Δ*pyrE2*, Δ*trpA*, Δ*rad25A*, Δ*rad25B*, Δ*rad25C*, Δ*rad25D*, *bgaHa*	
H1010	H999	Δ*pyrE2*, Δ*trpA*, *bgaHa*	

Table 2.3: Oligonucleotides

Name	Sequence (5′-3′)	Description
Chapter 3: DNA replication origins in *H. volcanii*		
F-1cdc	TCCCAGTTTTCCCGTAGATGAGG	Forward primer to amplify the intergenic region containing *oriC-2.*
R-1radh	TCGGATACCTCGGTTACGTCG	Reverse primer to amplify the intergenic region containing *oriC-2.*
RIP mapping primers: Cédric Norais (Norais *et al.,* 2007)		
OCN24 RIP-e	AGAGGAACTCATCGACTGCGTGGAG	Forward primer for RIP mapping of *ori-pHV1/4*
OCN29 RIP-f	TCGCAAGACAGTGTTGGAATCACG	Reverse primer for RIP mapping of *ori-pHV1/4*
OCN33 RIP-a	CGTAACTCGAACGGATACCGAGCAT	Forward primer for RIP mapping of replication origin *oriC-1*
OCN34 RIP-b	CGAATAACGATTTGGACGAATTCGG	Reverse primer for RIP mapping of replication origin *oriC-1*
OCN123 RIP-c	GACTCAGCACACACACCACGTTTC	Forward primer for RIP mapping of chromosomal replication origin *oriC-2*
OCN126 RIP-d	CCGGTGAAATACGATACAGGGCAGAC	Reverse primer for RIP mapping of replication origin *oriC-2*
OCN240 RIP-g	TGGTTGGGAGGTTGTGAACTGGAA	Forward primer for RIP mapping of replication origin *ori-pHV3*
OCN241 RIP-h	GTGTGTGTTCCCGAGGACAGATGG	Reverse primer for RIP mapping of replication origin *ori-pHV3*
OCN244 RIP-i	GGTCGAGACGTGTTCGGTGAGTGT	Forward primer for RIP mapping of replication origin *ori-pHV4-2*
OCN246 RIP-j	GGAACACACCCCCTCCACTTTGTC	Reverse primer for RIP mapping of replication origin *ori-pHV4-2*
Origin cloning: Cédric Norais (Norais *et al.,* 2007)		
OCN27	ACACCCTCATGTCGAGTGTGTCGAA	Forward primer for amplification of intergenic region upstream of *orc10*
OCN28	GGTCGGTAGAACGGTGCTGTTGACT	Reverse primer for amplification of intergenic region upstream of *orc10*
OCN35	GTGACTGAACGATGGTGGGTAGCC	Reverse primer for amplification of intergenic region upstream of *orc1*

OCN180	CGGGATCCGTGGTTGGGAGGTTGTGAAC	Forward primer for amplification of intergenic region upstream of *orc6*
OCN181	GGAATTCTGGACTCGCGTTTGTAACAG	Reverse primer for amplification of intergenic region upstream of *orc6*
OCN182	CGGGATCCATGTCCGCAGTTGTTCGTC	Forward primer for amplification of intergenic region upstream of *orc3*
OCN183	GGAATTCACTTAGGCGGGAACACACC	Reverse primer for amplification of intergenic region upstream of *orc3*
Chapter 4: *radA*-independent recombination		
Plas-R	TCACTCATTAGGCACCCCAGG	Plasmid-specific reverse primer to check for integration of pTA250 and pTA612.
PlasRC	CCTGGGGTGCCTAATGAGTGA	Plasmid-specific forward primer to check for integration of pTA441.
Chr-F	GCACATCTGCCCGAACAGG	Forward primer to check for pTA250 integration on pHV1.
oriC-2 chr-F	GAGGGGAAGATAGAGAACCGGAA	Forward primer to check for pTA612 integration on the main chromsome.
oriC-1 chrR2	CGAGAACCCGCGCGTATC	Reverse primer to check for pTA441 integration on the main chromosome.
8kb F2	CCCACTCGACACGGTGGTCAC	Forward primers to sequence 8 kb unidentified cloned band.
8kb F3	CCTGTGCGCGTATATGGAACTCCG	
8kb R2	GCCCATTAACTCTATAATGTCCC	Reverse primers to sequence 8 kb unidentified cloned band.
8kb R2a	GGGACATTATAGAGTTAATGGGC	
radARTF	GACGATACGCTTGTCGCCC	Forward internal primer for *radA*, used to check *radA* deletion.
radARTR	GTCCGCCTCCTCTGTGTTGAC	Reverse internal primer for *radA*, used to check *radA* deletion.
Chapter 5: Characterisation of *RAD25*-like genes		
rad25AF	CGAGTTTCTTCGTCGCTGTCC	Forward internal primer used to make a *rad25A* probe.
rad25AR	CAACGCCTTCGGTGACCAAC	Reverse internal primer used to make a *rad25A* probe.
rad25CF	CGATTTACAGGACCAGTGGATACG	Forward internal primer used to make a *rad25C* probe.

rad25CR	CCCCGAAAGCAGGATGCC	Reverse internal primer used to make a *rad25C* probe.
R25F	CGAGCGGGTCGTCCTCTTTC	Forward internal primer used to make a *rad25D* probe.
R25R	TGTCGGAGATGGGGGCGTAG	Reverse internal primer used to make a *rad25D* probe.
rad25AHindIII-F	CGACGGTATCGATAAGCTTCGATAG	Forward primer for upstream flanking region of *rad25A*, contains *Hin*dIII site. Used to generate *rad25A* deletion construct.
rad25AEcoRI-R	GTCGTCGAATTCGATTCGCATT	Reverse primer for upstream flanking region of *rad25A*, contains *Eco*RI site. Used to generate *rad25A* deletion construct.
rad25AEcoRI-F	CGAGAGGGGGTGAATTCGTAGTG	Forward primer for downstream flanking region of *rad25A*, contains *Eco*RI site. Used to generate *rad25A* deletion construct.
rad25AXbaI-R	GGTGGCGTCTAGAGGACTGCCT	Reverse primer for downstream flanking region of *rad25A*, contains *Xba*I site. Used to generate *rad25A* deletion construct.
R25EPF	CCGTGGGTACCCTCCCGGAACGAGCC	Forward primer for upstream flanking region of *rad25D*, contains *Kpn*I site. Used to generate *rad25D* deletion construct.
R25IPR	CATCGGATCCGAATACGGTCGCGACGG	Reverse primer for upstream flanking region of *rad25D*, contains *Bam*HI site. Used to generate *rad25D* deletion construct.
R25IPF2	ATTGGGATCCTCTTGACCGCTCTTAGC	Forward primer for upstream flanking region of *rad25D*, contains *Bam*HI site. Used to generate *rad25D* deletion construct.
R25EPR	CTGTCTAGACTGTCACGGTTCGTGTCCG	Reverse primer for upstream flanking region of *rad25D*, contains *Xba*I site. Used to generate *rad25D* deletion construct.
bgaBbF	CGAGCTCGGCACCATGGCTGAGACGT	Forward primer for insertion at *Bst*BI site in *bgaHa* to generate *bgaHa-Bb* allele, contains *Nco*I site.
bgaBbR	CGACGTCTCAGCCATGGTGCCGAGCT	Reverse primer for insertion at *Bst*BI site in *bgaHa* to generate *bgaHa-Bb* allele, contains *Nco*I site.

Miscellaneous plasmid construction		
hdrBsp	TGGCCTCATGAGCGGCGAGGAGC	Forward primer for subcloning *hdrB* coding sequence in pIL11 expression vector, contains *Bsp*HI site.
hdrXba	CTCCCACTCTCTAGAGTTACTCATCGG	Reverse primer for subcloning *hdrB* coding sequence in pIL11 expression vector, contains *Xba*I site.
hdrBXF	CGCGGTGTCTAGATGAGCGGCGAGGAGC	Forward primer for subcloning *hdrB* coding sequence downstream of *pyrE2* in pGB70 (pTA52) to generate *p.fdx*::*pyrE2*::*hdrB* construct, contains *Xba*I site.
hdrBHR	CCACTCGCCAAGCTTACTCATCGGATTCC	Reverse primer for subcloning *hdrB* coding sequence downstream of *pyrE2* in pGB70 (pTA52) to generate *p.fdx*::*pyrE2*::*hdrB* construct, contains *Hin*dIII site.

Table 2.4: Plasmids

Name/Alias	Notes
pMLH32	*H. alicantei bgaH* gene cloned in the shuttle vector pMDS20 (Holmes *et al.,* 1991), (Holmes and Dyall-Smith, 2000)
pCN11	pTA131 with 692 bp fragment of the intergenic region of *oriC-1*, amplified by PCR using OCN33 and OCN35 primers and inserted at *Not*I site. Constructed by Cédric Norais. (Norais *et al.,* 2007)
PCN12	pTA131 with 633 bp fragment of pTA194 containing the intergenic region of *ori-pHV1/4*, amplified by PCR using OCN27 and OCN28 primers and inserted at *Not*I site. Constructed by Cédric Norais. (Norais *et al.,* 2007)
pTA9	pBluescript KS with *H. volcanii* thymidylate synthase (*hts*) and hDHFR-2 (*hdrB*) 3.56 kb *Mbo*I-*Hin*dIII chromosomal fragment inserted at *Bam*HI-*Hin*dIII sites. (Ortenberg *et al.,* 2000)
pTA22/ pMDS99	*H. volcanii* shuttle vector derived from pMDS17 (*ori-pHV2* from pWL102, and pOK12 plasmid), with Mev^R (*hmgA*) from *H. hispanica*. (Wendoloski *et al.,* 2001)
pTA29/pSJS1140	*H. volcanii radA* ORF on *Kpn*I-*Sph*I fragment cloned in pUC118. (Sandler *et al.,* 1996; Woods and Dyall-Smith, 1997)

pTA 35/ pBluescript II SK+	Standard *E. coli* vector.
pTA49	pBluescript II SK+ with *H. volcanii* 3.7 kb *Sau*3AI chromosomal fragment containing *trpA*, inserted at *Bam*HI site. (Allers *et al.,* 2004)
pTA52/pGB70/pIL11	*H. volcanii pyrE2* ORF under *H. salinarum fdx* (ferredoxin) promoter in pUC19 (Bitan-Banin *et al.,* 2003)
pTA80	pTA52 with 2.44 kb *Kpn*I-*Hin*dIII Δ*radA* fragment (*Hin*dIII blunt) from pMDS41 (Woods and Dyall-Smith, 1997), inserted at *Kpn*I and *Eco*RI (blunt). Constructed by Thorsten Allers.
pTA95	pTA52 with 1.39 kb *Xba*I Δ*trpA* fragment from pTA92 (Allers *et al.,* 2004), inserted at *Xba*I. Constructed by Thorsten Allers.
pTA102	pTA52 with 3.43 kb *Hin*dIII-*Xba*I *bgaH* fragment from pMLH32 inserted at *Hin*dIII and *Xba*I. Constructed by Thorsten Allers.
pTA103	pBluescript I with fill-in of *Asp*718 (*Kpn*I) site. Constructed by Thorsten Allers.
pTA107	pBluescript II SK+ with 1.9 kb *Sty*I fragment (blunt-ended) of pTA44 (Allers *et al.,* 2004), containing *leuB*, inserted at *Eco*RV site. Constructed by Thorsten Allers.
pTA128	pTA103 with 3.37 kb *Hin*dIII-*Age*I chromosomal fragment from H54 containing *bgaHa*, inserted at *Hin*dIII and *Xma*I sites. Constructed by Thorsten Allers.
pTA131	pBluescript II SK+ with 0.7 kb *Bam*HI/*Xba*I (both blunt-ended) *p.fdx*::*pyrE2* fragment from pTA52 inserted at *Psi*I site. (Allers *et al.,* 2004)
pTA144	Insertion of *bgaBb* oligos (*bgaBbF*, *bgaBbR*, containing *Nco*I site) at *Bst*BI site of p128, to produce *bgaHa-Bb* allele in pTA103 (pBluescript I with fill-in of *Kpn*I site). Constructed by Thorsten Allers.
pTA151	pTA131 with insertion of 3.5 kb *Hin*dIII-*Xba*I fragment from pTA144 containing *bgaHa-Bb* allele. Constructed by Hien-Ping Ngo.
pTA155	pTA131 with insertion of 0.8 kb Δ*hdrB* PCR construct, made from pTA9 using internal primers with *Nhe*I sites and external primers with *Hin*dIII/*Xba*I sites, inserted at *Hin*dIII/*Xba*I site. (Allers *et al.,* 2004)
pTA160	pTA131 with insertion of 0.8 kb Δ*hdrB* PCR construct, made from pTA9 using internal primers with *Nhe*I sites and external primers with *Hin*dIII/*Xba*I sites (same primers used in pTA160 construction), inserted at *Hin*dIII/*Xba*I site. Constructed by Hien-Ping Ngo.
pTA176/pUC19	Standard *E. coli* vector.

pTA187	*H. volcanii hdrB* ORF (*Bsp*HI-*Xba*I PCR fragment, hdrXba and hdrBsp), in pTA52 between *Nco*I and *Xba*I sites, under ferredoxin (*fdx*) promoter of *H. salinarium*. Constructed by Thorsten Allers.
pTA194	pTA131 containing 5 kb *Hpa*II genomic fragment from WR340 with *ori-pHV1/4* inserted at *Cla*I site. (Norais *et al.,* 2007)
pTA230	pTA131 with insertion of 3.75 kb *Nco*I-*Hin*dIII (both blunt-ended) fragment containing *ori-pHV2* from pWL-Nov, at *Pci*I site (blunt-ended). (Allers *et al.,* 2004)
pTA250	pTA131 containing 1.1 kb *Aci*I fragment of pTA194 insert with intergenic region of *ori-pHV1/4* inserted at *Cla*I site. (Norais *et al.,* 2007)
pTA252	pTA131 containing 3 kb *Bam*HI-*Stu*I fragment of pTA194 with *ori-pHV1/4* inserted at *Bam*HI and *Not*I sites. (Norais *et al.,* 2007)
pTA266	pTA252 with deletion of 1 kb *Bsm*I-*Xba*I fragment containing *ori-pHV1/4*, replaced by 1 kb *Bam*HI-*Xba*I fragment of pTA106 containing *trpA* marker (Norais *et al.,* 2007)
pTA274	pTA230 shuttle vector with insertion of 3 kb *Sca*I-*Hin*dIII *bgaHa* fragment from pTA128 at *Asp*7I8-*Not*I sites (all sites except *Sca*I blunt-ended by Klenow). Constructed by Thorsten Allers.
pTA298	*H. volcanii trpA* ORF (*Pci*I-*Sph*I PCR fragment amplified from pTA49) in pIL11/pTA52 between *Nco*I and *Sph*I sites, i.e. under ferredoxin (*fdx*) promoter of *H. salinarium*. Constructed by Thorsten Allers.
pTA313	pTA131 containing 2.2 kb *Aci*I genomic fragment from H230 with chromosomal replication origin *oriC-1* inserted at *Cla*I site. (Norais *et al.,* 2007)
pTA314	pTA131 with *H. volcanii* 3.96 kb *Aci*I chromosomal *oriC-1* fragment (partial digest) of pTA194, inserted at *Cla*I site. Constructed by Thorsten Allers.
pTA315	pTA131 with *H. volcanii* 5 kb *Aci*I chromosomal *oriC-1* fragment (partial digest) of pTA194, inserted at *Cla*I site. Constructed by Thorsten Allers.
pTA323	pTA29 with 0.965 kb *Bam*HI (blunt-ended) *p.fdx*::*trpA* fragment from pTA298 inserted between *Nco*I site and *Not*I site in *radA* (sites blunt-ended before ligation) to give Δ*radA*::*trpA* deletion. Constructed by Thorsten Allers.
pTA324	pTA131 with 2.09 kb *Eco*RI Δ*radA*::[*p.fdx*::*trpA*] insert from pTA323, inserted at *Eco*RI site. Constructed by Thorsten Allers.

pTA333	pUC19 with H1 8.5 kb *Sac*I-*Nsp*I chromosomal fragment containing putative homing endonuclease, inserted at *Sac*I-*Sph*I sites. Constructed by Thorsten Allers.
pTA354	*H. volcanii* shuttle vector with *pyrE2* marker derived from pTA131. Contains 948 bp *Bmg*BI-*Eco*RV fragment of pTA250 with *ori-pHV1/4* inserted at *Pci*I site. (Norais *et al.,* 2007)
pTA368	pTA354 shuttle vector with with 1.4 kb *Not*I (blunt-ended) Mev^R fragment from pMDS99 inserted at *Sap*I site (blunt-ended). Constructed by Thorsten Allers.
pTA381	pTA368 shuttle vector with 1.6 kb *Kpn*I-*Bst*BI fragment containing *radA* ORF from pTA29, inserted at *Kpn*I-*Cla*I sites. Constructed by Thorsten Allers.
pTA396	pTA52 with insertion of 685 bp *Xba*I-*Hin*dIII *hdrB* PCR product (hdrBHR, hdrBXF) at *Xba*I-*Hin*dIII sites directly downstream of *pyrE2*. Constructed by Thorsten Allers.
pTA401	pTA35 with 1.4 kb *Eco*RI (klenow blunted) fragment from pTA22 containing Mev^R marker inserted at *Psi*I site. Constructed by Sam Haldenby.
pTA402	pBluescript II SK+ with 1.25 kb *Bam*HI/*Hin*dIII (both blunt-ended) *p.fdx*::*pyrE2*::*hdrB* fragment from pTA396 inserted at *Psi*I site. Constructed by Thorsten Allers.
pTA409	Shuttle vector based on pTA402 with 948 bp *Bmg*BI-*Eco*RV fragment from pTA250 (containing *ori-pHV1/4*) inserted at klenow-blunted *Pci*I site. Constructed by Thorsten Allers.
pTA411	pTA409 shuttle vector with 1.7 kb *Kpn*I-*Bam*HI fragment containing *radA* gene from pTA381, inserted at *Kpn*I-*Bam*HI sites. Constructed by Thorsten Allers.
pTA416	pBluescript II containing 9.3 kb *Not*I genomic DNA fragment from WR340 with chromosomal origin *oriC-2* inserted at *Not*I site. (Norais *et al.,* 2007)
pTA419	pTA131 with 5.5 kb *Nhe*I-*Eco*RI fragment of pTA416 containing *oriC-2*, *rad25D* and *orc5*, inserted at *Xba*I and *Eco*RI sites. Constructed by Thorsten Allers.
pTA441	pTA131 containing 1.1 kb *Sau*3AI-*Hind*III fragment of pTA313 with intergenic region of *oriC-1* inserted at *Bam*HI and *Hind*III sites. (Norais *et al.,* 2007)
pTA451	pTA131 with 4.75 kb *Xho*I-*Hin*dIII fragment of pTA333, containing *orc4* inserted at *Xho*I and *Hin*dIII sites. Constructed by Thorsten Allers.

pTA453	pTA131 with insertion of 1.5kb Δ*rad25D* PCR construct, made from pTA416 using internal primers with *Bam*HI sites (R25IPR, R25IPF2) and external primers with *Kpn*I and *Xba*I sites (R25EPF, R25EPR), inserted at *Kpn*I and *Xba*I sites. Constructed by Zhenhong Duan.
pTA506	pTA131 with insertion of 2.6 kb *Hin*dIII-*Bsr*GI (both blunt) *bgaHa* fragment from pTA128 at *Asp*718/*Not*I sites (both blunt). Constructed by Thorsten Allers.
pTA531	pTA419 with replacement of 856 bp *Rsr*II (blunt) *oriC-2* DUE fragment by 965 bp *Bam*HI (blunt) *trpA*$^{+}$ fragment from pTA298. Constructed by Thorsten Allers.
pTA532	pTA441 with replacement of 470 bp *Eco*RI-*Bsp*EI (both blunt) *oriC-1* DUE fragment by 965 bp *Bam*HI (blunt) *trpA*$^{+}$ fragment from pTA298. Constructed by Thorsten Allers.
pTA611	pTA131 with insertion of 1.7 kb intergenic region adjacent to *orc4* made from pTA452 digested with *Apa*I/*Bgl*II, inserted at *Apa*I and *Bam*HI sites. (Norais *et al.,* 2007)
pTA612	pTA131 containing 2.4 kb fragment of pTA416 with intergenic region of *oriC-2*, amplified by PCR using F-1cdc and R-1radh primers and inserted at *Eco*RV site. (Norais *et al.,* 2007)
pTA613	pTA131 with insertion of 1.2 kb *leuB*Δ*Bb* construct, made from pTA107 digested with *Bst*BI/*Eco*RI, inserted at *Cla*I and *Eco*RI sites. (Norais *et al.,* 2007)
pTA617	pTA506 with deletion of 2 kb *Eco*RV-*Bam*HI (blunt) *bgaHa* fragment, to generate Δ*bgaHa* construct. Constructed by Thorsten Allers.
pTA628	pTA131 with insertion of H98 *Cla*I 6.7 kb fragment (containing DNA polymerase IV (X family) and Sulfatase arylsulfatase A-like protein) at the *Cla*I site.
pTA631	pTA230 with insertion of 1.4 kb *Not*I fragment from pTA22 with Mev^{R} marker, at the *Not*I site. Constructed by Thorsten Allers.
pTA636	pTA631 shuttle vector with insertion of 1.7 kb *Kpn*I-*Bam*HI *radA* fragment of pTA411, at *Kpn*I/*Bam*HI sites. Constructed by Thorsten Allers.
pTA641	pTA354 shuttle vector with *bgaHa* gene, generated by replacement of 784 bp *Nco*I-*Bst*XI fragment of pTA354 with 3.7 kb *Nco*I-*Bst*XI fragment of pTA274. Constructed by Thorsten Allers.
pTA647	pTA131 with insertion of *Psh*AI fragment of pTA628 containing the *PolX* intergenic region, inserted in to the *Eco*RV site.

pTA725	pBluescript II SK+ with H98 4.6 kb *Eco*RI chromosomal fragment containing *rad25B*, inserted in to the *Eco*RI site.
pTA731	pBluescript II SK+ with H98 5 kb *Kpn*I/*Xba*I chromosomal fragment containing *rad25C*, inserted in to *Kpn*I/*Xba*I sites.
pTA732	Δ*rad25C* construct derived from pTA731 by deletion of the 1 kb *Cla*I-*Bst*BI *rad25C* fragment.
pTA733	Δ*rad25B* construct derived from pTA725 by deletion of the 1.6 kb *Nru*I *rad25B* fragment.
pTA734/pCN26	pTA131 containing 693 bp fragment of the intergenic region of pHV3 replication origin, amplified by PCR using OCN180 and OCN181 primers and inserted at *Bam*HI and *Eco*RI sites. (Norais *et al.,* 2007)
pTA735/pCN27	pTA131 containing 593 bp fragment of the intergenic region of contig 452 replication origin, amplified by PCR using OCN182 and OCN183 primers and inserted at *Bam*HI and *Eco*RI sites. (Norais *et al.,* 2007)
pTA737	2.9 kb deletion construct for *rad25B* from pTA733 cut with *Eco*RI/*Not*I and *Bsp*HI (to aid band extraction) inserted at *Eco*RI/*Not*I sites of pTA131.
pTA738	3.9 kb deletion construct for *rad25C* from pTA732 cut with *Xba*I/*Kpn*I inserted at *Xba*I/*Kpn*I sites of pTA131.
pTA765	pTA401 with insertion of the 2.4 kb *Eco*RI/*Hin*dIII *oriC-2* fragment from pTA612 at the *Eco*RI/*Hin*dIII sites.
pTA767	pTA401 with insertion of the 1.1 kb *Eco*RI/*Xho*I *ori-pHV1/4* fragment from pTA250 at the *Eco*RI/*Xho*I sites.
pTA787	pTA401 with insertion of the 700 bp *Eco*RI/*Not*I *ori-pHV3* fragment from pTA734 at the *Eco*RI/*Not*I sites.
pTA788	pTA401 with insertion of the 600 bp *Eco*RI/*Not*I *ori-pHV4-2* fragment from pTA735 at the *Eco*RI/*Not*I sites.
pTA812	pTA401 with insertion of the 1.7 kb *Bam*HI/*Xho*I intergenic region adjacent to *orc4* from pTA451 at the *Bam*HI/*Xho*I sites.
pTA821	p401 with insertion of the 800 bp *Eag*I *oriC-1* sequence from pTA313 at the *Not*I site.
pTA878	pBluescript II SK+ with 5.2 kb *Bam*HI/*Hin*dIII H98 genomic fragment containing *rad25A*, inserted at *Bam*HI/*Hin*dIII sites.
pTA882	pTA131 with insertion of *Xba*I/*Hin*dIII Δ*rad25A* PCR construct generated from pTA878 using internal primers with *Eco*RI sites (rad25AEcoRI-F, rad25AEcoRI-R,) and external primers with *Hin*dIII and *Xba*I sites (rad25AHindIII-F, rad25AXbaI-R), inserted at *Xba*I/*Hin*dIII sites.

pTA894	pBluescript II SK+ with 8 kb H884 *Eco*RI fragment inserted at the *Eco*RI site.
pTA946	pTA441 with replacement of 470 bp *Eco*RI-*Bsp*EI (blunted) *oriC-1* DUE fragment by 716 bp *Bam*HI/*Xba*I (blunted) *hdrB* fragment from pTA187.
pTA1045	pTA641 with insertion of 774 bp *bgaHa Eco*RI fragment from pTA151 at the *Eco*RI site.
pTA1100	pTA131 with H1023 4.7 kb *Aci*I chromosomal *oriC-3* fragment, inserted at *Cla*I site.

2.1.2 *Haloferax volcanii* media

Sterilised by autoclaving (1 minute 121°C) unless otherwise stated. Plates stored in sealed bags at 4°C and dried for at least 30 minutes before use.

30% SW: 240 g NaCl, 30 g $MgCl_2 \cdot 6H_2O$, 35 g $MgSO_4 \cdot 7H_2O$, 7 g KCl, 20 ml 1 M Tris.HCl pH 7.5, dH_2O to 1 litre (not autoclaved)

18% SW: 200 ml 30% SW, 100ml dH_2O, 2 ml $CaCl_2$ added after autoclaving.

Trace elements: 36 mg $MnCl_2.4H_2O$, 44 mg $ZnSO_4.7H_2O$, 230 mg $FeSO_4.7H_2O$, $CuSO_4.5H_2O$, dH_2O to 100 ml

Hv-Min salts: 30 ml 1M NH_4Cl, 36 ml 0.5 M $CaCl_2$, 6 ml trace elements

Hv-Min carbon source: 41.7 ml 60% DL-lactic acid sodium salt (Sigma), 37.5 g succinic acid Na_2 salt.$6H_2O$ (Sigma), 3.15 ml 80% glycerol, dH_2O to 250 ml, pH 6.5

10X YPC: 50 g Yeast extract (Difco), 10 g Peptone (Oxoid), 10 g Casamino acids, 17.6 ml 1 M KOH, dH_2O to 1 L (not autoclaved, used immediately)

10X Ca: 50 g Casamino acids, 23.5 ml 1M KOH, dH_2O to 1 L (not autoclaved, used immediately)

Hv-YPC agar: (Contains uracil, tryptophan and leucine) 5 g Agar (Bacto), 100 ml dH_2O, 200ml 30% SW, 33 ml 10X YPC, 2 ml 0.5 M $CaCl_2$. Microwaved without 10X YPC to dissolve agar. 10X YPC added, then autoclaved. $CaCl_2$ added prior to pouring.

Hv-Ca agar: (Contains leucine) 5 g Agar (Bacto), 100 ml dH_2O, 200 ml 30% SW, 33 ml 10X Ca, 2 ml 0.5 M $CaCl_2$, 36 µg biotin (Sigma), 288 µg thiamine (Sigma). Microwaved without 10X Ca to dissolve agar. 10X Ca added, then autoclaved. $CaCl_2$, thiamine and biotin added prior to pouring.

Hv-YPC broth: (Contains uracil, tryptophan and leucine) 100 ml dH_2O, 200 ml 30% SW, 33 ml 10X YPC, 2 ml 0.5 M $CaCl_2$. Autoclaved and $CaCl_2$ added when cool.

Hv-Ca^+ broth: (Contains leucine) 150 ml 30% SW, 7.5 ml 1 M Tris.HCl pH 7.5, 6.5 ml Hv-Min carbon source, 3 ml Hv-Min Salts, 500 µl 0.5 M KPO_4 buffer (pH 7.5), 250 µl 36 µg biotin (Sigma), 288 µg thiamine (Sigma), 25 ml 10X Ca. 30% SW autoclaved. All additive added once cooled.

High SW: 280 g NaCl, 30 g $MgCl_2 \cdot 6H_2O$, 35 g $MgSO_4 \cdot 7H_2O$, 7 g KCl, 20 ml 1 M Tris.HCl pH 7.5, dH_2O to 1 L (not autoclaved)

Low SW: 200 g NaCl, 30 g $MgCl_2 \cdot 6H_2O$, 35 g $MgSO_4 \cdot 7H_2O$, 7 g KCl, 20 ml 1 M Tris.HCl pH 7.5, dH_2O to 1 L (not autoclaved)

Hv-high/low salt media: 200 ml of warmed high/low SW added to 5 g agar dissolved in 100 ml dH_2O. 33 ml of 10X YPC or Ca added before autoclaving. $CaCl_2$ added prior to pouring.

Table 2.5: *Haloferax volcanii* media supplements
All media supplements were supplied by Sigma. All solutions were sterilised by filtration through a 0.2 µm filter (except for mevinolin).

Supplement	Abbreviation	Final concentration	Stock concentration
Leucine	+ Leu	50 µg/ml	10 mg/ml
Uracil	+Ura	50 µg/ml	50 mg/ml
Thymidine	+Thy	50 µg/ml (+50 µg/ml hypoxanthine in Hv-Ca)	4 mg/ml
Tryptophan	+Trp	50 µg/ml	10 mg/ml
5-FOA	+5-FOA	50 µg/ml (+10 µg/ml uracil)	50 mg/ml
Mevinolin	+Mev	4 ug/ml	10 mg/ml

2.1.3 *Haloferax volcanii* buffers and solutions

All solutions were sterilised by filtration through a 0.45 µm filter.

Alkaline Spheroplasting Solution: 1 M NaCl, 27 mM KCl, 50 mM Tris.HCl pH 8.5, 15% sucrose. Also known as Buffered Spheroplasting Solution.

Neutral Spheroplasting Solution: 1 M NaCl, 27 mM KCl, 15% sucrose, pH 7.5. Also known as Unbuffered Spheroplasting Solution.

Spheroplast Dilution Solution: 23% SW, 15% sucrose, 37.5 mM $CaCl_2$

Regeneration Solution: 18% SW, 1X YPC, 15% sucrose, 30 mM $CaCl_2$

Transformant Dilution Solution: 18% SW, 15% sucrose, 30 mM $CaCl_2$

For high and low salt regeneration solution/transformant dilution solution the high/low SW was diluted 2:1in dH_2O and used instead of 18% SW.

ST Buffer: 1 M NaCl, 20 mM Tris.HCl pH 7.5

Lysis Solution: 100 mM EDTA pH 8, 0.2% SDS

Plug lysis solution: 1% sarkosyl, 500 mM EDTA, 20 mM Tris-HCl pH 8.8

Plug wash buffer: 100 mM EDTA, 25 mM Tris-HCl pH 7.5

2.1.4 *Escherichia coli* media

Sterilised by autoclaving.

LB: 10 g tryptone (Bacto), 5 g yeast extract (Difco), 10 g NaCl, 10 g agar (Bacto) when required for LB agar, dH_2O to 1 L

SOC: 20 g tryptone (Bacto), 5 g yeast extract (Difco), 0.58 g NaCl, 0.186 g KCl, 2.03 g $MgCl_2$, 2.46 g $MgSO_4$, dH_2O to 1 L

Table 2.6: *Escherichia coli* media supplements

Supplement	Abbreviation	Final concentration	Stock concentration
Ampicillin	+Amp	50 µg/ml	4 mg/ml
Tetracycline	+Tet	3.5 µg/ml	1 mg/ml
Isopropyl-β-D-thiogalactoside[a]	IPTG	29 µg/ml	8 mg/ml
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside[a]	X-Gal	36 µg/ml	10 mg/ml

[a]BlueTech solution (Mirador)

2.1.5 Other buffers and solutions

TE:10 mM Tris.HCl pH 8.0, 1 mM EDTA

TBE: 89 mM Tris, 89 mM boric acid, 2 mM EDTA

TAE: 40 mM Tris.HCl, 20mM acetic acid, 1 mM EDTA

Denaturing Solution: 1.5 M NaCl, 0.5 N NaOH

20X SSPE: 525.9g NaCl, 82.8g NaH_2PO_4, 28.2g EDTA in 3 L dH_2O, pH 7.4

100X Denhardt's Solution: 2% Ficoll 400, 2% PVP (poly-vinyl pyrrolidone) 360, 2% BSA (bovine serum albumin, Fraction V)

Prehybridisation Solution: 24 ml dH_2O, 12 ml 20X SSPE, 2 ml 20% SDS, 2 ml 100X Denhardt's solution

Hybridisation Solution: 1.5 g dextran sulphate, 9 ml 20X SSPE, 1.5 ml 20% SDS, 18 ml dH_2O

Low Stringency Wash Solution: 2X SSPE, 0.5% SDS

High Stringency Wash Solution: 0.2X SSPE, 0.5% SDS

Neutralising Buffer: 1.5 M NaCl, 0.5 M Tris.HCl, 1 mM EDTA

Gel Loading Dye (5X): 50 mM Tris.HCl, 100 mM EDTA, 15% Ficoll (w/v), 0.25% Bromophenol Blue (w/v), 0.25% Xylene Cyanol FF (w/v)

2.2 Methods

2.2.1 Nucleic acid manipulation

Plasmid DNA extraction

QIAGEN QIAprep Spin Miniprep or Plasmid Midi Kits were used to obtain circular plasmid DNA from *E. coli* strains by alkaline lysis. Plasmid purifications were carried out according to the manufacturer's guidelines using 2 ml of cell culture for Miniprep kits and 300 ml for Midi Kits. For diagnostic digests of plasmid DNA requiring a small yield plasmid purifications were carried out using the Miniprep kit. The Plasmid Midi Kit was used when large, ultra-pure DNA yields were required for frozen stocks of DNA and transformations of *H. volcanii*. Plasmid DNA was eluted in 30 µl buffer EB (QIAgen) when using the Miniprep kits and resuspended in 200 µl TE after ethanol precipitation for Midi-preps. Plasmid DNA samples were frozen at -20°C.

Nucleic acid purification

PCR, Klenow-end filled DNA, ligation, restriction digest and dephosphorylated DNA products were all purified using the QIAquick PCR Purification kit (QIAgen). DNA was extracted from agarose gels using the QIAquick Gel Extraction kit (QIAgen). These kits function by salt and pH dependent adsorption of DNA to a silica-gel membrane in order to separate the desired nucleic acid from small oligonucleotides and proteins. DNA was eluted in 30 µl buffer EB (QIAgen).

For high quality DNA preparation (library ligations) or DNA concentration/resuspension, ethanol precipitation was used. 0.1 volumes of 3 M sodium acetate (pH 5.3) and 2 volumes of cold 100% ethanol were added to DNA samples and incubated at -20°C for 1 hour. Samples were centrifuged at 4°C, 14,000 ×*g* for 30 minutes. The supernatant was removed and pellets were washed in 1 ml 70% ethanol followed by centrifugation at 4°C, 14,000 ×*g* for 30 minutes. The supernatant was removed and pellets were air-dried before resuspension in sterile H_2O or TE.

Restriction digests

All restriction enzymes were supplied by New England Biolabs (NEB) and were carried out according to the manufacturers instructions (NEB). NEB buffers were

chosen so that no digest was carried out in conditions where any enzyme had activity <75%. If no suitable buffer existed for a given double digest sequential digests were performed. The first restriction digest was purified and the 30 µl elution was used for the secondary digestion.

Dephosphorylation of vector DNA

To prevent self-ligation of vector DNA, 5′ phosphate groups were removed following restriction digest. Samples were incubated with 5 units of Antarctic phosphatase (NEB) and 1X Antarctic phosphatase buffer (NEB) for 15 minutes at 37°C.

Klenow end filling

When DNA ends produced by restriction digests were incompatible for ligation, DNA Polymerase I large (Klenow) fragment (NEB) was used to generate blunt ends from 5′ overhangs. 1 µg DNA samples were incubated with 33 µM dNTPs, 1X Buffer 2 (NEB) and 0.5 units of Klenow for 30 minutes at room temperature.

Ligation

Ligations were performed using T4 DNA ligase (NEB). Reactions contained 400 U ligase and 1X T4 ligase buffer. Ligations were carried out at 15°C for 16 hrs. For vector:insert ligations, reactions contained a molar ratio of ~3:1 insert to vector DNA. Ligations were either ethanol precipitated and resuspended in 5 µl H_2O or purified using the QIAquick PCR Purification kit (QIAgen). In either case 5 µl was used to transform *E. coli* (XL1-Blue MRF′).

DNA sequencing and oligonucleotide synthesis

Synthesis of oligonucleotides, all sequencing reactions and their analysis, was carried out by the Biopolymer Synthesis and Analysis Unit, University of Nottingham.

Nucleic acid quantification

The concentration and purity of plasmid preparations was measured using the 260 nm, 280 nm absorbance ratio from spectrophotometer measurements (Beckman Coulter DU 530). For plasmids used in quantitative *H. volcanii* transformations the purity and concentration was measured using a Nanodrop 1000 (Thermo Scientific) and verified by electrophoresis.

PCR amplification

Amplification of DNA was performed using either Phusion or DyNAzyme EXT, (Finnzymes). Phusion was used for amplifications where higher fidelity was required. Dynazyme was used for diagnostic amplifications and those using genomic DNA as the template. For DyNAzyme EXT, 200 µM of each dNTP, 0.5 µM of each primer, 10-50 ng of template DNA, 1X Optimised DyNAzyme Buffer, 5% DMSO and 1 U of Dynazyme were used in each reaction. For Phusion, 200 µM of each dNTP, 0.5 µM of each primer, 10 ng of template DNA, 1X Phusion GC buffer, 5% DMSO and 1 U of Phusion were used in each reaction. 50 µl reactions were used when the product was required for downstream applications and 20 µl reactions for purely diagnostic reactions. Reaction conditions are shown in Table 2.3 and annealing temperatures (Tm°C) were calculated using the nearest neighbour method (Breslauer *et al.,* 1986). All reactions were carried out using a Techne TC-512 thermocycler.

Table 2.7: PCR reaction conditions

Step	DyNAzyme	Phusion
Initial Denaturation	94°C, 120 s	98°C, 30 s
Denaturation	94°C, 15 s	98°C, 10 s
Annealing	Tm°C, 20 s	Tm°C, 15 s
Extension	72°C, 50 s/kb	72°C, 20 s/kb
Final Extension	300 s	300 s

2.2.2 Agarose gel electrophoresis

Basic

Agarose gels for separation of small DNA fragments (100-600 bp) were cast using RESponse Research PCR agarose. Agarose gels for separation of larger DNA fragments were cast using AGTC Bioproducts Ltd hi-res standard agarose powder. TBE buffer was used for the majority of gels, TAE buffer was used for 16 hour gels. Analytic gels were supplemented with ethidium bromide (final concentration 0.5 µg/ml). Gels for size-based DNA extraction were post-stained with ethidium bromide after removal of sample lanes. Visualisation was carried out using a UV transilluminator (UVP inc.) and sample lanes were protected from UV exposure using foil while the appropriate band was excised. DNA samples were mixed with a

1/5 by final volume of 5X gel loading dye. Electrophoresis was carried out at 5 V/cm.

Pulsed-field gel electrophoresis

2 ml agarose plug DNA preparations were loaded on to a 1% agarose 0.5X TBE gel. Electrophoresis was performed at 14°C in a CHEF mapper (Bio-Rad) using 0.5X TBE buffer, a voltage gradient of 6 V/cm, a switch angle of 120° and switch times of 0.47 s (initial) to 1 min 33.83 s (final). Total run time was 20 hours 18 minutes.

Two-dimensional agarose gel electrophoresiss

30 ml plug DNA preparations were incubated for 1 hour in 10 ml wash solution with 0.5 mM PMSF at 37°C, 90 oscillations/minute. Plugs were then washed in 10 mM Tris at 37°C, 90 oscillations/minute for 1 hour. Equilibration took place in the required NEB buffer for 1 X 3 hours and 2 X 1 hour periods at 37°C, 90 oscillations/minute. Plugs were sliced in half and melted in 1.5 ml eppendorfs at 70°C for 10 minutes and cooled to 37°C in a hot block. 1 µl of *Nru*I (50000 units/ml) was added and digestion was performed overnight at 37°C in a hot block (Eppendorf Thermomixer Comfort) shaking at 300 rpm. A further 25 units of *Nru*I were added to the molton plugs and incubated as before for 4 hours. A 0.4% Seakem gold agarose TBE gel was made with plexiglass dividers positioned to contain the wells and the first dimension DNA migration. Plugs were melted at 70°C for 15 minutes and 50 µl was loaded in each well. After 5 minutes to allow the plugs to set, the first dimension was run for 20 hours at 12 V at room temperature. The first dimension run was stained in 250 ml 1X TBE with 10 µl ethidium bromide (10 mg/ml). Well-containing strips were arranged perpendicular to the DNA migration direction on a 15 cm × 25 cm gel tray. A 200 ml 1% Seakem low melting agarose gel in 1X TBE with 8 µl ethidium bromide (10 mg/ml) was poured around the gel strips to submerge them. The second gel dimension was run at 4°C at 130V for 12 hours with 1X TBE circulating buffer containing 0.3 µg/ml ethidium bromide.

2.2.3 *Escherichia coli* microbiology

Growth of *Escherichia coli* and storage

Cultures on solid media were grown at 37°C in a static incubator (LEEC). Liquid cultures were grown in a ground shaker at 37°C with 60 rpm rotation. For short-term storage, plates were stored at 4°C. For long term storage, 25% (v/v) glycerol was added to cultures which were flash frozen on dry ice and stored at -80°C.

Preparation of electrocompetent *Escherichia coli* cells

An overnight culture of XL1-Blue or N2338 was grown. Antibiotic selection was dependant on the strain: XL1-Blue cultures were supplemented with 3.5 µg/ml tetracycline and N2338 cultures were grown in LB. Cells were diluted 1/100 in LB broth supplemented with the appropriate antibiotics and grown at 37°C to an OD_{650} of 0.5-0.8. Cells were pelleted at 6000 ×*g* for 12 minutes at 4°C, the supernatant was removed and the pellet resuspended in an equal volume of ice-cold sterile 1 mM HEPES (pH 7.5). This process was repeated using 0.5 volumes 1 mM HEPES, 0.25 volumes 1 mM HEPES + 10% glycerol, 0.1 volumes 1mM HEPES + 10% glycerol, and finally 0.001 volumes 1mM HEPES + 10% glycerol. Cells were flash frozen on dry ice and stored in 50 µl aliquots at -80°C.

Transformation of *Escherichia coli* by electroporation

5 µl of a ligation reaction or a plasmid was added to 40 µl of electrocompetent cells while on ice. The mixed sample was added to a pre-chilled electroporation cuvette (1 mm electrode gap, GENEFLOW). The cuvette was placed in an *E. coli* gene pulser (BioRad) and subjected to a 1.8 kV pulse. 1 ml of SOC was added immediately and samples were incubated at 37°C with 250 rpm rotation for 1 hour. Cells were plated onto selective LB plates and incubated at 37°C overnight.

Generation of unmethylated (*dam*$^-$) plasmid DNA

H. volcanii contains a restriction endonuclease that cleaves DNA at 5′- GA$^{(CH3)}$TC-3′ sequences. Therefore unmethylated DNA is required for efficient *H. volcanii* transformation (Holmes *et al.,* 1991). All plasmids were passaged through the *dam*$^-$ *dcm*$^-$ *E. coli* strain N2338, which is deficient in GATC methylation, before being used to transform *H. volcanii*.

2.2.4 *Haloferax volcanii* microbiology

Growth of *Haloferax volcanii* and storage

Cultures on solid media were grown at 45°C in a static incubator (LEEC) in a plastic bag to prevent desiccation. Liquid cultures were grown in the same incubator with 8 rpm rotation. Cultures on solid media were stored at room temperature for short-term storage. For long-term storage, 20% glycerol (v/v) was added to cultures, mixed and flash frozen on dry ice. Frozen cultures were stored at -80°C.

Transformation of *Haloferax volcanii*/ARS assay

1-4 colonies were used to inoculate 10 ml Hv-YPC broth which was incubated at 45°C overnight (~16 hours). Δ*radA* strains were inoculated using several colonies and incubated for ~21 hours. 20 ml of culture was used for Δ*radA* strain transformations. When the OD_{650} ≈0.6-0.8, cells were pelleted in a round bottom tube at 3300 ×*g* for 8 minutes. Cells were gently resuspended in 2 ml buffered spheroplasting solution, spun as before and resuspended in 600 μl buffered spheroplasting solution. For each transformation 200 μl of cells were transferred to a new round bottom tube. 20 μl 0.5 M EDTA, pH 8.0 was added to the side of the tube, mixed by gentle inversion and incubated at room temperature for 10 minutes. Transforming DNA was set up in a 30 μl volume: 5 μl 0.5 M EDTA, pH 8.0, 15 μl unbufferred spheroplasting solution, 10 μl *dam*$^-$ DNA (~1-2 μg), and added in the same manner as EDTA. For quantitative co-transformations 0.5 μg of each plasmid was used, standard quantitative transformations used 1 μg. After 5 minutes at room temperature 250 μl 60% PEG solution was added and mixed by shaking the tube ~10 times gently horizontally. After a 30 minute incubation at room temperature, 1.5 ml spheroplast dilution solution was added and mixed by tube inversion. Cells were pelleted at 3300 ×*g* for 8 minutes and the supernatant was removed. 1 ml regeneration solution was added (+ thymidine if required) and a wide-bore tip used to transfer the intact cell pellet to a 4 ml sterile tube. After undisturbed incubation at 45°C for 90 minutes the pellet was resuspended by tapping the tube and then returned to 45°C for 3 hours with 8 rpm rotation. Cells were transferred to a round bottom tube and pelleted at 3300 ×*g* for 8 minutes. The supernatant was removed and the cells resuspended gently in 1 ml of transformant dilution solution. Dilutions were made using transformant dilution solution and 100 μl of was plated out on selective

media. For transformations at low and high salt concentrations regeneration solution, transformant dilution solution and plates with the appropriate NaCl concentration were used.

Quantitative estimation of ARS plasmid stability

H112 containing pTA250 or pTA441 (maintained by selection on Hv-Ca agar and in Hv-Ca broth) was used to inoculate a 5 ml of Hv-YPC broth. When the cultures reached OD_{650} 0.5, a 1 ml aliquot was transferred to a round bottom tube and pelleted at 3300 ×*g* for 8 minutes. Cells were resuspended and diluted in 18% SW and 100 µl plated on Hv-YPC agar. These colonies were patched on to Hv-Ca to determine the fraction of uracil$^+$ colonies that indicated plasmid presence. The OD_{650} 0.5 cultures were diluted 1/250 in to fresh Hv-YPC broth. This was repeated three times and plasmid loss was calculated using the formula:

$l = 100(1 - \sqrt[n]{ura^+})$ l = % plasmid loss/generation

n = number of generations

ura^+ = fraction of uracil$^+$ cells (will grow on Hv-Ca)

2.2.5 *Haloferax volcanii* phenotypic assays

Ultraviolet (UV) radiation sensitivity

1 colony was used to inoculate 5 ml Hv-YPC broth which was incubated at 45°C overnight. When the OD_{650} ≈0.5, cells were diluted in 18% SW (10^{-1}-10^{-8}). Duplicate 20 µl spots of each dilution were spotted on to Hv-YPC plates and allowed to dry before UV exposure (1 J/m^2/second, 254 nm peak). Plates were incubated at 45°C in a black bin bag.

Mitomycin C (MMC) sensitivity

1 colony was used to inoculate 5 ml Hv-YPC broth which was incubated at 45°C overnight. When the OD_{650} ≈0.5, cells were diluted in 18% SW (10^{-1}-10^{-8}) Duplicate 20 µl spots of each dilution were spotted on Hv-YPC agar containing 0.02 or 0.04 µg/ml MMC. Plates were allowed to dry before incubation.

Methanesulfonic acid methyl ester/Methyl methanesulfonate (MMS) sensitivity

1 colony was used to inoculate 5 ml Hv-YPC broth which was incubated at 45°C overnight. When the OD_{650} ≈0.5, cells were divided in to 1 ml aliquots in screw cap tubes. A 4% solution of MMS was prepared in 18% SW, this stock solution was

further diluted in 18% SW to give final concentrations of 0.04-0.08%. 20 µl of the MMS dilutions were spotted on to the lid caps, these were then replaced and the tubes inverted. After 1 hour incubation at 45°C, 450 rpm (Eppendorf Thermomixer Comfort), cells were diluted in 18% SW (10^{-2}-10^{-6}) and duplicate 20 µl spots were spotted on Hv-YPC plates. Plates were allowed to dry before incubation.

Hydrogen peroxide (H_2O_2) sensitivity

1 colony was used to inoculate 5 ml Hv-YPC broth which was incubated at 45°C overnight. When the OD_{650} ≈0.5, cells were divided in to 1 ml aliquots in screw cap tubes. H_2O_2 was made fresh for each experiment by dilution in 18% SW (300mM/100mM). 20 µl of the H_2O_2 dilutions were spotted on to the lid caps, these were then replaced and the tubes inverted. After 1 hour incubation at 45°C, 450 rpm (Eppendorf Thermomixer Comfort), cells were diluted in 18% SW (10^{-2}-10^{-6}) and duplicate 20 µl spots were spotted on Hv-YPC plates. Plates were allowed to dry before incubation.

Growth curves

1 colony was used to inoculate 5 ml Hv-YPC broth which was incubated at 45°C overnight. When the OD_{650} ≈0.5, cells were diluted 1/10000 in Hv-YPC broth. At 2 hour time intervals during incubation the OD_{650} was measured and an aliquot taken. This sample was diluted in 18% SW (10^{-1}-10^{-8}) and duplicate 20 µl spots spotted on Hv-YPC plates. Growth curves were followed until the OD_{650} >0.9. Generation time was calculated using the formula:

$$G = t/n$$

$$n = \frac{\log b - \log B}{\log 2}$$

G = generation time

t = time

n = number of generations

b = ending cell number

B = starting cell number

Time period chosen using the exponential growth phase OD_{650} ≈0.1-0.6.

Competition assay

5 ml Hv-YPC broth was innoculated with 1 colony of a *bgaHa*$^+$ strain and another culture was started with a wild-type strain. When the OD_{650} ≈0.5, cells were plated on Hv-YPC. Each strain was diluted 1/1000 and 100 µl of each was added to 9.8 ml of Hv-YPC broth. The mixed culture was grown until OD_{650} ≈0.5 when it was

diluted 1/100 in to fresh Hv-YPC broth This was repeated three times and before each dilution cells were plated on to Hv-YPC. Once colonies were visible they were sprayed with an X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) solution (BlueTech, Mirador) and incubated overnight at 45°C.

2.2.6 Construction of *Haloferax volcanii* gene deletion strains

Genomic gene clone

The candidate gene in *H. volcanii* was found in the TIGR genome sequence and unique restriction sites at least 500 bp upstream and downstream were identified. Restriction sites were chosen that were also present in the multiple cloning site of pBluescript II SK+ to facilitate cloning. Genomic DNA was isolated from strain H26 and digested with the appropriate enzymes for 8 hours. Agarose gel electrophoresis was performed on the digested genomic DNA (0.8% TAE gel, 16 hours at 50 V) and DNA was purified from a gel slice containing the predicted fragment size +/- 0.5 kb. Purified DNA was ligated to cut, dephosphorylated pBluescript II SK+. This plasmid library DNA was used to transform XL1-Blue MRF′ *E. coli*. Transformants were selected on LB +Amp and several colonies were tested for presence of the genomic clone by plasmid isolation and diagnostic digest. If no plasmids contained the clone the original transformants were patched onto fresh plates and incubated overnight at 37°C. PCR primers were designed to amplify a 150-500 bp region of the desired gene to generate a probe for Southern hybridisation. Colony lifts (p82) were performed from the patched plates and used in a Southern hybridisation with the PCR product acting as the template for a radiolabelled probe. Colonies that hybridised with the probe contained the desired gene and this was confirmed by plasmid isolation followed by diagnostic digests and DNA sequencing.

Deletion constructs

Restriction digest

If restriction sites were available that cut the genomic clone so that self-ligation resulted in deletion (complete or majority of the gene, for example see 5.3) of the gene they were used. This construct was then subcloned in to pTA131, which contains a selectable marker (*pyrE2*) for use in *H. volcanii*.

PCR

If no suitable restriction sites were found then primers were designed to enable amplification and ligation of the gene flanking regions. Two pairs of primers were designed, one pair to amplify a region >150 bp upstream of the promoter of the gene to be deleted and the other pair to amplify >150 bp downstream of the termination codon of the gene to be deleted. For the two internal primers, the same novel restriction site was introduced so that both PCR products could be ligated. The external primers included restriction sites present in the polylinker region of pTA131. A maximum of 3 bp were changed to introduce these novel restriction sites. Any site that was introduced was unique and not present anywhere else in the PCR products. Phusion DNA polymerase was used to ensure maximum fidelity.

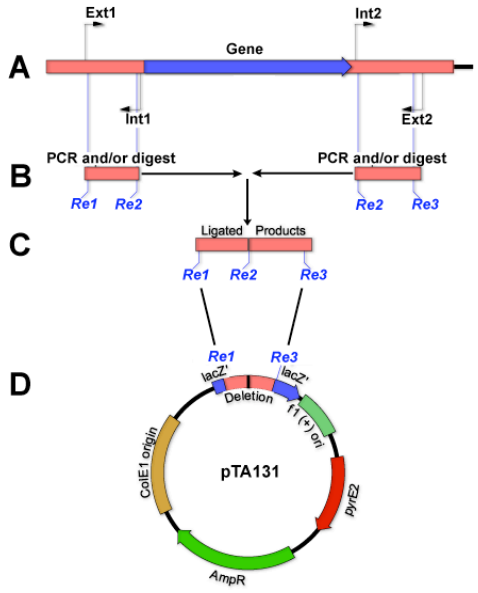


Figure 2.1: Schematic of gene deletion construct generation

(A) Restriction sites flanking the gene are digested from the genomic clone and ligated *or* pairs of primers are designed to amplify flanking sequences of the gene (Int1 and Ext1, and Int2 and Ext2) and to contain novel restriction sites (Re1, Re2 and Re3).

(B) PCR amplification products are digested with Re2 and ligated.

(C) Ligated product is digested with Re1 and Re3, and ligated in to the polylinker of pTA131.

(D) Plasmid constructed for gene deletion.

Gene deletion

Deletion constructs were integrated at the native gene locus by transforming *H. volcanii* with the *dam*$^-$ version of the deletion construct cloned in pTA131. By relieving selection for uracil, intramolecular recombination is encouraged and the integrated plasmid and native gene were lost from the chromosome. Intramolecular recombinants that have lost the plasmid were counter selected using 5-fluoroorotic acid, which is converted to toxic 5-fluorouracil in *ura*$^+$ cells but not *ura*$^-$ cells. Depending upon the location of the recombination event either wild-type or deletion mutants can be recovered. Replacing the gene with a selectable marker such as *trpA*$^+$ enables direct selection for deletion mutants.

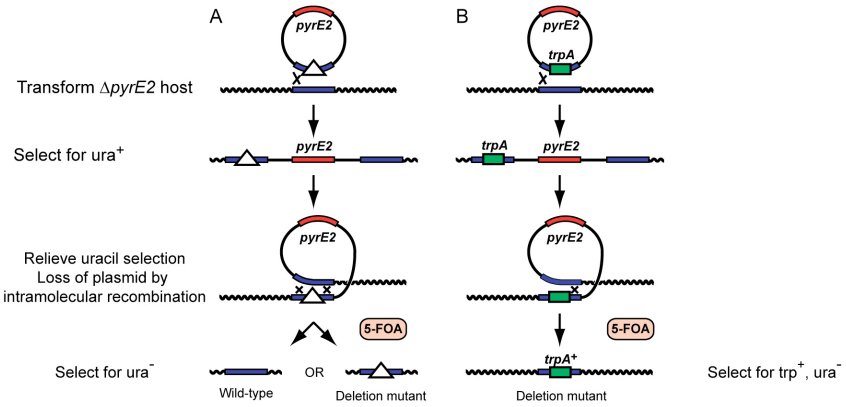


Figure 2.2: *Haloferax volcanii* gene deletion strategies
(A) Pop-in pop-out gene deletion method.
(B) Pop-in pop-out gene replacement method. Adapted from Allers and Mevarech, 2005.

A Δ*pyrE2 H. volcanii* strain was transformed with a non-replicating plasmid containing the deletion construct. Transformants were plated on selective media and to confirm integration of the plasmid at the correct locus restriction digest of genomic DNA was performed. The restriction digest was chosen so that the orientation of integration could be determined. Electrophoresis of digested genomic DNA was performed and the DNA transferred on to a membrane (p82). Southern hybridisation was performed using a probe made by excising the deletion construct from pTA131.

After verifying the genotype of the integrant strain, a second recombination event to remove the integrated *pyrE2*-encoding plasmid from the chromosome was encouraged by relieving uracil selection. 5 ml of Hv-YPC broth was inoculated and incubated at 45°C with 8 rpm rotation. At late-log phase (OD_{650} >0.8) cultures were

diluted 1/500 in fresh culture and grown again. This serial dilution was repeated twice more and then 1 ml of culture at OD_{650} ~0.6 was pelleted at 3300 ×*g* for 8 mins, the cell pellet was resuspended in 18% SW. Appropriate dilutions were made (10^0, 10^{-1} and 10^{-2}) and plated on Hv-Ca +5FOA, in addition to other required additives. If the deletion was marked with *trpA* selection was maintained for *trpA*$^+$.

Pop-out plates were incubated at 45°C for at least five days and colonies were patched on to Hv-YPC plates. If *radA* was being deleted then the pop-out plates were left for 7-10 days and only small colonies were patched because Δ*radA* strains have a significant growth defect. Colony lifts from the patched plates were analysed by Southern hybridisation using part of the deleted gene as a probe. Strains deleted for the gene could be identified as they did not hybridise with the probe. The genotype of these putative deletion candidates was checked using restriction digestion of genomic DNA and Southern analysis.

Genes that were difficult to delete (e.g. *radA*) were complemented with an episomal copy of the gene during the pop-out phase by maintaining selection for an episomal plasmid. After serial overnight cultures to facilitate pop-out of the gene, selection for the plasmid was relieved and two more serial overnights cultured. Plating on 5-FOA media then proceeded as normal, this selected against the integrated deletion construct plasmid and any episomal plasmid still present because the episomal plasmid is also marked with *ura*$^+$.

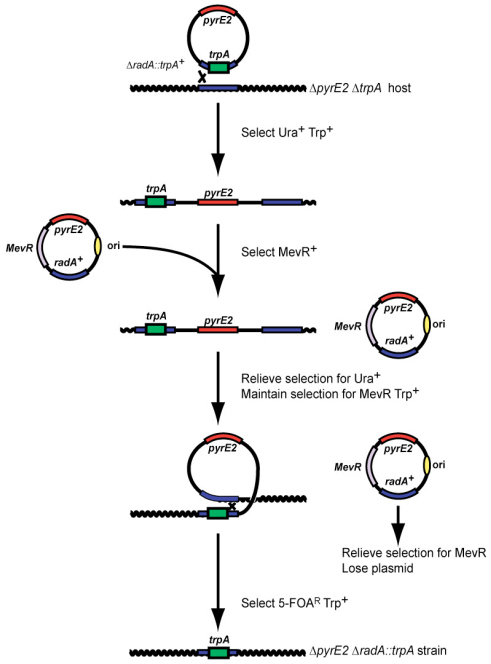


Figure 2.3: *Haloferax volcanii* gene deletion with complementing plasmid
Strategy for deleting genes using complementation during the pop-out stage.

2.2.7 *Haloferax volcanii* DNA isolation

Intact DNA preparation in agarose plugs

2 ml of culture (OD_{650} ~1) was pelleted at 3300 ×*g* for 10 min at 4°C and resuspended in 1 ml of cold buffered spheroplasting solution + 0.1% NaN_3, and pelleted again. Cells were gently resuspended in 80 µl of buffered spheroplasting solution and transferred to 42°C. After 2 minutes at 42°C cells were mixed with 100 µl of 1.5% low-melt agarose (in 0.5X buffered spheroplasting solution, 100 mM EDTA) and pipetted into plug moulds (Bio-Rad). The following plug incubation steps took place with agitation (90 oscillations/minute). Plugs were incubated in 5 ml of plug lysis solution with proteinase K (0.5 mg/ml) for 3 hours at 52°C and were then transferred to fresh plug lysis solution with proteinase K (0.5 mg/ml) and RNaseA (30 mg/ml) and incubated overnight at 52°C. Plugs were washed three times in 10 ml of plug wash buffer at 37°C, equilibrated in 0.5X TBE for 90 min at 20°C and exposed to 50 Gy of γ radiation (137Cs, 375 Gy/s) to linearize circular DNA molecules.

High quality genomic DNA preparation

1 ml of culture (OD_{650} ≈0.6-0.8) was pelleted in a round bottom tube at 3300 ×*g* for 5 minutes. The supernatant was removed and cells were resuspended gently in 200 µl of ST buffer. 200 µl of lysis solution was then added and the tube inverted. 1 ml of 100% ethanol was overlayed and DNA at the interface was spooled on to a capillary tip. The sample was spooled until the liquid was homogenous and clear. Spooled DNA was washed twice in 100% ethanol and air-dried. The DNA was resuspended in 500 µl TE and then isopropanol precipitated by adding 400 µl isopropanol and 50 µl sodium acetate. After pelleting at 3300 ×*g* for 5 minutes the pellet was washed with 70% ethanol. DNA was resuspended in 100 µl TE and incubated at 45°C, 500 rpm (Eppendorf Thermomixer Comfort) for 1 hour with 1 µl of RNase (30 mg/ml). DNA was left at 4°C overnight to completely resuspend and was stored at 4°C.

Rough genomic DNA/total DNA preparation for Southern analysis

1 ml of culture (OD_{650} >0.6) was pelleted for 1 minute at 16000 ×*g*. The supernatant was removed and the pellet resuspended in 400 µl of dH_2O to lyse the cells. DNA was incubated at 70°C for 10 minutes and stored at 4°C for no more than a month before use. 5 µl was used for digest and Southern analysis.

Rough genomic DNA/total preparation for PCR templates/Southern analysis

A sterile toothpick was used to transfer cells from a patch or a colony in to 100 µl of dH_2O for 5 minutes, known as the "stab method". 5 µl was used for digest and Southern analysis.

Phenol-chloroform DNA purification

1 ml of culture (OD_{650} >0.6) was pelleted for 1 minute at 16000 ×*g*. The supernatant was removed and the pellet resuspended in 200 µl of dH_2O to lyse the cells. 200 µl of phenol was added and the mixture incubated at 65°C for 1 hour. The mixture was pelleted for 5 minutes at 16000 ×*g* and the aqueous phase was transferred to a fresh tube. 400 µl of cold 100% ethanol was then added and the sample was pelleted for 10 minutes at 16000 ×*g* and washed in 1 ml of 70% ethanol. DNA was resuspended in 100 µl of TE.

2.2.8 *Haloferax volcanii* ARS library construction

To isolate ARS in a genetic screen, genomic DNA was prepared using a "genome maxiprep" protocol. 50 ml of *H. volcanii* culture of OD_{650} ≈0.6-0.8) was pelleted at 6000 ×*g* for 5 minutes. Cells were resuspended in 5 ml ST buffer before adding 5 ml of lysis buffer. After inversion, 25 ml of 100% ethanol was added and DNA at the interface was spooled on to a 1 ml glass pipette. DNA was washed in 100% ethanol and air dried before resuspension in 15 ml TE. To concentrate DNA 1.5 ml 3 M sodium acetate pH5.2 and 12 ml isopropanol was added. DNA was transferred in to a 70% wash and air dried before resuspension in 5 ml TE. 10 µl of RNase (30 mg/ml) was added and DNA was incubated at 45°C >1 hour.

DNA was partially digested using sub-optimal buffer for 30 minutes and separated by size using electrophoresis. Fragments of the desired size were purified from the agarose and ligated in to pTA131. The ligation was purified by ethanol precipitation and used to transform N2338. Transformants were pooled using a spreader to collect colonies within LB broth. The plasmid library was purified using the QIAprep Plasmid Midi Kit and 1 µg was used to transform *H. volcanii* to prototrophy for uracil.

2.2.9 Genotype analysis

Vacuum transfer of DNA to nylon membrane

200 ml 0.8% TAE gels for Southern analysis were run overnight at 50 V with buffer circulation and stained in 0.5 μg/ml ethidium bromide for 30 minutes. Gel-embedded DNA was acid-nicked for 10 minutes in 0.25 M HCl, washed for 10 minutes in dH_2O and denatured in denaturing solution for 45 minutes. A 15 × 25 cm Bio-Rad Zeta-Probe GT positively charged membrane was soaked in dH_2O before 5 minutes of equilibration in denaturing solution. The membrane and gel were positioned on a Vacugene XL gel blotter (Pharmacia Biotech) and covered in denaturing solution (gel on top). Transfer took place for 1 hour at 40 mbar using a Vacugene Pump (Pharmacia Biotech). After transfer the membrane was briefly (30 seconds) washed in 2X SSPE and then air-dried on Whatman paper. The membrane was UV-crosslinked using 120 mJ/cm^2.

Colony lift

Candidate colonies and suitable controls were patched on to plates using sterile sticks and incubated until patches had grown. Bio-Rad Zeta-Probe GT membrane was left on the plate surface for 1 minute. Using forceps the filter was transferred colony side up to Whatman paper soaked in 10% SDS for 3 minutes. The filter was then transferred to Whatman paper soaked in denaturing solution for 5 minutes. Finally the filter was placed on Whatman paper soaked in neutralising solution for 3 minutes. Neutralisation was repeated and the filter was washed briefly (30 seconds) in 2X SSPE before being air-dried on Whatman paper. The membrane was UV-crosslinked using 120 mJ/cm^2.

2.2.10 Southern hybridisation

Membranes from colony lifts or transfer of agarose gels were prehybridised for 3 hours at 65°C in 40 ml prehybridisation solution with 800 μl of denatured fish DNA (10 mg/ml Roche). Before addition of the radiolabelled probe the prehybridisation solution was replaced with 30 ml hybridisation solution. 1 μl of 1 μg/ml 1 kb ladder, 4 μl of the template DNA and 10 μl of dH_2O was denatured at 100°C for 5 minutes and then chilled on ice. 4 μl of High Prime random priming mix (Roche) and 0.74Mbq [α–^{32}P] dCTP (GE Healthcare, Amersham Redivue), was added and incubated at 37°C for 30 minutes. The radiolabelled probe was purified using a

BioRad P30 column and then mixed with 450 µl of fish sperm DNA (10 mg/ml Roche). The probe was denatured at 100°C for 5 minutes and chilled on ice before hybridisation. Membranes were hybridised overnight at 65°C and then washed with 50 ml low stringency solution for 10 minutes at 65°C. After 10 minutes, the wash solution was replaced with 50 ml fresh low stringency wash solution for 30 minutes. This second wash was replaced with 50 ml of high stringency wash solution for 30 minutes, this was then repeated. After the final wash, the membrane was allowed to dry briefly on Whatman paper and wrapped in Saran wrap. The membrane was exposed to a phosphorimager screen (Fujifilm BAS Cassette2 2325) for at least 12 hours. The screen was scanned using a Molecular Dynamics STORM 840 scanner and the resulting graphical file was refined using L-Process and analysed using Image Quant (Fujifilm).

2.2.11 Flow cytometry

Cultures of DS2 were set up using various amounts of inoculum in Hv-YPC broth. The next day various dilutions were made from a range of OD_{650} and these were incubated for a further 24 hours. On day three every 3 hours 450 µl of culture was spun down at 3003 ×*g*, 4°C for 8 minutes and resuspended in 450 µl 18% SW. 50 µl of acridine orange (0.1 mg/ml) was added and the sample run on an Apogee A40 flow cytometer with settings of: sample flow 0.7 µl /minute, volume 110 µl , sheath pressure = 150. PMT voltaged were 450 V for forward light scatter (LS1) and 500 V for emission at 510-580 nm (FL1). Data was analysed using Flow Jo (Tree Star Inc.).

2.2.12 Microscopy

Acridine orange (0.1 mg/ml) was diluted 1:100 in filtered 18% SW. A thin layer of 1% agarose gel in filtered 18% SW was spread on a glass slide and allowed to dry. 45 µl of cells (OD_{650} <0.5) were mixed with 5 µl of the diluted acridine orange. The tube was agitated for 20 seconds to mix and then left at room temperature for 3 minutes. 5 µl was spotted on to the agarose and left to dry. Cells were visualised using a BX-52 Olympus microscope equipped with a 100× immersion lens and a coolSNAP^{TM}HQ camera (Photometrics), using the JP4-CFP-YFP filterset 86002v2 (Chroma). Images were processed using MetaMorph and Adobe photoshop.

2.3 Replication initiation point mapping

All replication initiation point (RIP) mapping assays were carried out by Cédric Norais. *H. volcanii* DS2 was diluted to an OD_{650} of 0.15 in 100 ml of Hv-YPC media, grown to OD_{650} nm of 0.3, and pelleted. Cells were resuspended in 4 ml of lysis buffer (25 mM Tris-HCl [pH 7.5], 20 mM EDTA, 100 mM NaCl, 200 μg/ml proteinase K, 1% SDS). After 1 hour incubation at 50 °C, 4 g of CsCl and 100 μl of Hoechst-33342 (5 mg/ml) were added and the refractive index adjusted to 1.410 with CsCl (1 g/ml). DNA was purified by CsCl gradient ultracentrifugation. To enrich for replicating intermediates, total DNA was passed down a BND-cellulose column pre-equilibrated with NET buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, and 1 M NaCl) to selectively bind single-stranded DNA. After washing with NET buffer, bound DNA was eluted with NET buffer with 1.8% caffeine at 50 °C. DNA was isopropanol-precipitated and resuspended in TE buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA) at 1 μg/μl. After phosphorylation of 5′-OH ends with T4 polynucleotide kinase (Promega, http://www.promega.com), DNA was treated with λ-exonuclease to digest 5′ nicked DNA ends; replication intermediates protected by RNA primers are unaffected by this treatment. Primer extension reactions used ~500 ng of enriched replicating intermediates, 25 ng of radiolabeled primer (labeled using [c32P]-ATP, and T4 polynucleotide kinase) and 2 units of Deep Vent (exo-) DNA polymerase. After 30 cycles of reaction (60 s at 94 °C, 60 s at 70 °C, and 90 s at 72 °C), amplification products were separated on a 6% polyacrylamide gel under denaturing conditions. Radioactive material was detected using a Phosphorimager system (Amersham, www.gehealthcare.com). Control experiments using linearised plasmid DNA isolated from *E. coli* were performed under similar conditions.

2.4 Bioinformatics

DNA curvature was calculated using a BEND.IT server (http://hydra.icgeb.trieste.it/;kristian/dna/bend_it.html), which predicts in qualitative terms a curvature propensity of a given DNA sequence using DNase I based bendability parameters and the consensus bendability scale.

pI was calculated using the Henderson-Haselbach equations which are based on calculating the percentage disassociation for each amino acid in a protein at different pHs. MacVector 10.0.2 was used for these calculations.

Nucleotide representation disparities were calculated using either ORIGINX (www.cbs.dtu.dk/services/GenomeAtlas/suppl/origin) or ZPLOTTER (http://tubic.tju.edu.cn/zcurve) programs. Publicly available TIGR *H. volcanii* genome sequences released in June 2006 were used to calculate local minima and maxima in nucleotide disparities. The resulting data from each program was plotted using Origin Pro 7.5 software (OriginLab Corporation, www.originlab.com).

Chapter 3: DNA replication origins in *Haloferax volcanii*

3.1 Replication genes

The core replisome proteins are all encoded by the *H. volcanii* genome. With one exception (see Table 3.1), these genes are all located on the main chromosome (Hartman *et al.*, in preparation). In common with other archaea, the *H. volcanii* set of replication genes are more closely related to the eukaryotic rather than the bacterial homologues. Analysis of the preliminary *H. volcanii* genome sequence showed that it contains 16 homologues of the putative initiator factor *cdc6/orc1* across the four replicons. When referring to specific *H. volcanii cdc6/orc1* homologues the *orc* nomenclature will be used in order to remain consistent with published work (Norais *et al.*, 2007).

Table 3.1: *cdc6/orc1* homologues in *Haloferax volcanii*
Result of BLAST search querying the *H. volcanii* genome sequence (TIGR) with the COG1474: Cdc6-related protein, AAA superfamily ATPase consensus sequence (see Appendix 7.1).

Gene	Replicon	Contig coordinates	BLAST score	Expect value
orc1	Chromosome	258-1952	315	2×10^{-87}
orc2	Chromosome	569203-570429	231	4×10^{-62}
orc3	pHV4	152-1402	230	8×10^{-62}
orc4	Chromosome	1888354-1889583	218	4×10^{-58}
orc5	Chromosome	1594707-1595912	216	2×10^{-57}
orc6	pHV3	201-1433	213	1×10^{-56}
orc7	pHV4	258668-257391	207	9×10^{-55}
orc8	pHV1	56369-57616	200	1×10^{-52}
orc9	Chromosome	174892-176016	186	2×10^{-48}
orc10	pHV1	101-1327	172	3×10^{-44}
orc11	Chromosome	2162076-2162873	160	1×10^{-40}
orc12	pHV4	66289-65261	155	3×10^{-39}
orc13	pHV4	56897-55455	154	9×10^{-39}
orc14	Chromosome	2161873-2160851	144	6×10^{-36}
orc15	Chromosome	1998477-1999643	67	1×10^{-12}
orc16	Chromosome	1403626-1404828	55	5×10^{-9}

The 16 *cdc6/orc1* genes are spread across the replicons of *H. volcanii*. *orc11* is missing the C-terminal winged helix domain that is required for DNA binding (Singleton *et al.,* 2004). COG (clusters of orthologous groups) consensus sequences represent the current state of sequence databases and are therefore fluid. The BLAST scores and E-values in Table 3.1 differ from previously published values (Norais *et al.,* 2007), this is due to the use of a revised consensus sequence for COG1474. The consensus sequence used here also identified two additional *cdc6/orc1* homologues on the main chromosome that were named *orc15* and *orc16* based on their BLAST scores.

The majority of identified archaeal replication origins are adjacent to *orc1*/*cdc6* genes so these results provided a starting point for the identification of replication origins in *H. volcanii.* The large number of *cdc6/orc1* homologues is similar to that found in other halophilic archaea. For example 10 *cdc6/orc1* genes are found in *Halobacterium* sp. NRC-1 across three replicons (Berquist and DasSarma, 2003) and 17 are present in *H. marismortui* over nine replicons (Baliga *et al.,* 2004). However, not all *cdc6/orc1* genes in archaea are associated with replication origins so it is unlikely that all 16 *cdc6/orc1* genes in *H. volcanii* are adjacent to active origins of DNA replication.

3.2 Genetic screen for autonomously replicating sequences

Given the multitude of potential replication origin sites in *H. volcanii*, a genetic screen for autonomously replicating sequences (ARS) was carried out to avoid bias. The initial genetic screens, strain construction and cloning described in 3.2 were carried out by Thorsten Allers (University of Nottingham, Nottingham) and the repeat analysis and RIP mapping by Cédric Norais (Université Paris-Sud, Orsay). Since this work is an important foundation of this thesis it will be described here. Several figures in this chapter have been adapted from Norais *et al.,* 2007, this paper is provided as a reference in appendix 7.4.

Genomic DNA from *H. volcanii* strain WR340 (Bitan-Banin *et al.,* 2003) was partially digested with the rare-cutter *Hpa*II. A library of DNA fragments 4-8 kb in size was cloned in to the non-replicating *pyrE2*-marked plasmid pTA131 at the compatible *Cla*I site (Allers *et al.,* 2004). This was used to transform *H. volcanii* strain H49 (Δ*pyrE2*, Δ*radA*) (Bitan-Banin *et al.,* 2003) to prototrophy for uracil

(*pyrE2*$^{+}$). RadA is the archaeal recombinase and *H. volcanii radA* deletion strains are defective in homologous recombination (Woods and Dyall-Smith, 1997). Using this strain for the ARS assays means that there will be no false positives due to integration of the plasmid in to the genome, any colonies will be the result of a library insert acting as an origin of replication. *H. volcanii* plasmid preparations from seven of the thirty-five transformants were passaged through *E. coli* so that high quality plasmid DNA could be recovered for sequencing. In each case the insert corresponded to the same sequence. To ensure that this was not a technical artefact of the screen, the library construction was repeated using *Aci*I to partially digest *H. volcanii* genomic DNA. This library was also cloned in to pTA131 at the compatible *Cla*I site and used to transform H49 (Δ*pyrE2*, Δ*radA*). Plasmid DNA from six *Aci*I library transformants was sequenced and each contained the same region of the *H. volcanii* genome isolated in the initial screen.

Figure 3.1 shows the region isolated in the genetic screens for ARS elements. It includes two divergently transcribed genes, *orc10* and a hypothetical protein, separated by a 910 bp intergenic region featuring a 104 bp A/T-rich putative duplex unwinding element (DUE, 68% A/T vs. 33% for the entire genome) and several direct repeats. These features are similar to those found at other characterised archaeal origins of DNA replication. However, the nucleotide sequence of the similar direct repeats (Figure 3.1, panel B) shows only a weak resemblance to the origin recognition box (ORB). ORB sequences are conserved at many archaeal origins of replication and have been shown to bind Cdc6/Orc1 proteins (Robinson *et al.,* 2004).

The intergenic region of the cloned insert was used to probe a Southern blot of intact *H. volcanii* DNA displayed on a pulsed-field gel (PFG) to verify the location of this replication origin (Figure 3.1, panel C). Two bands were observed which correspond in size to replicons pHV4 (690 kb) and pHV1 (86 kb) (Charlebois *et al.,* 1991). The intensity of the two bands is similar which suggests that this replication origin is present on both pHV1 and pHV4 and that both replicons are present in all cells. However, in the genome sequence this replication origin could only be found on pHV1. Since the location of this replication origin was ambiguous it was named *ori-pHV1/4*.

To determine the minimal *ori-pHV1/4* sequence needed for DNA replication activity, an ARS subclone library was constructed. A partial *Aci*I digest was carried out on the

5 kb insert of the *ori-pHV1/4* ARS clone pTA194. Fragments of 1-1.3 kb, 1.3-1.6 kb, 1.6-2 kb, 2-3 kb and 3-4 kb were purified separately and cloned in to the compatible *Cla*I site of pTA131 (*pyrE2*). Transformants of *H. volcanii* H49 (Δ*pyrE2*, Δ*radA*) were obtained with DNA from each size range. Plasmid DNA from six transformants in the smallest size range (1-1.3 kb) was sequenced and in all cases the insert included the 910 bp intergenic region. The minimal origin was further delimited using PCR to amplify a 633 bp fragment of the intergenic region (Figure 3.1, panel A, pCN12). This fragment could maintain the plasmid in an ARS assay (pCN12 was constructed by Cédric Norais).

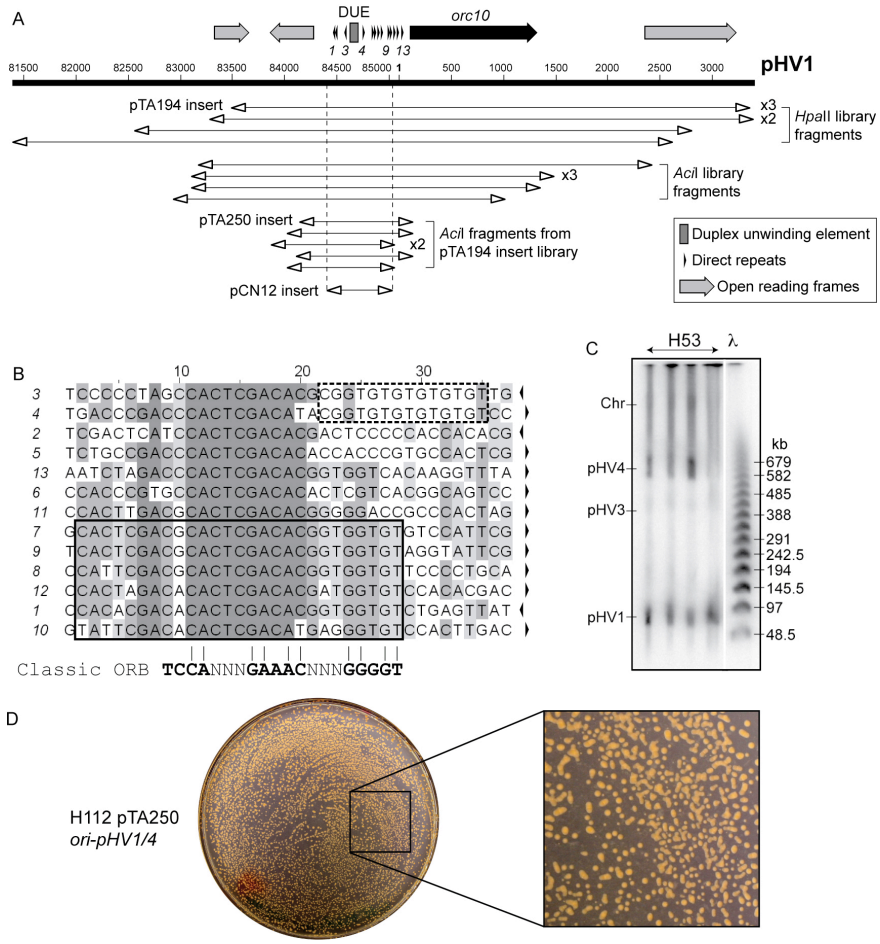


Figure 3.1: Features of *ori-pHV1/4*

(A) Sequence features of *ori-pHV1/4* and the regions used in ARS assays.

(B) Sequence of repeats (numbered in A) within the intergenic region adjacent to *orc10*. Boxed sequences are complete ORB elements, arrows indicate repeat orientation, conserved sequences are shaded and a sequence motif in repeats surrounding the DUE is indicated by a dashed box. Analysis by Cédric Norais.

(C) Southern blot of intact *H. volcanii* DNA from strain H53, probed with the intergenic *ori-pHV1/4* locus (pTA252 *Xba*I/*Bsm*I).

(D) H112 (Δ*radA*, Δ*pyrE2*) transformed with 1 µg of plasmid pTA250 (*ori-pHV1/4*). 100 µl of a 100-fold dilution plated on Hv-Ca and incubated at 45°C for 15 days.

3.2.1 Deletion of *ori-pHV1/4*

H. volcanii has more than one replicon so it was notable that only a single ARS element was recovered from the genetic screen. *ori-pHV1/4* seemed to dominate the ARS library so in order to isolate ARS elements that might correspond to other origins the *ori-pHV1/4* locus was deleted.

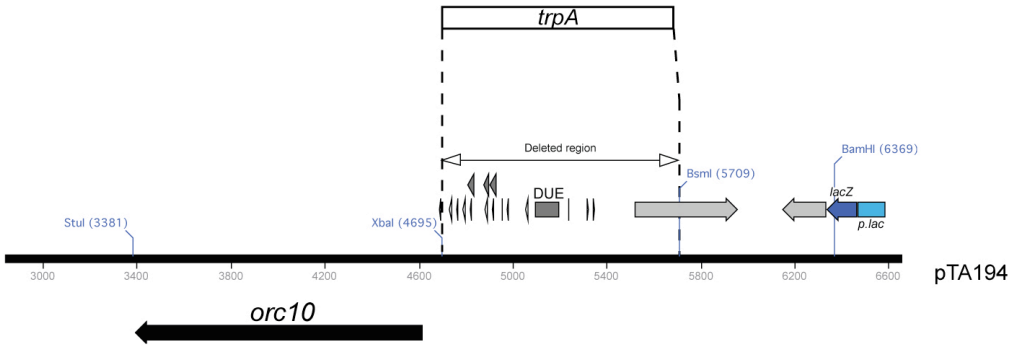


Figure 3.2: Generation of *ori-pHV1/4* deletion construct
The *ori-pHV1/4* locus. The region containing the DUE and repeats replaced by the *trpA* marker in the *ori-pHV1/4* deletion construct are highlighted.

The *Bam*HI/*Stu*I fragment of pTA194 containing *ori-pHV1/4* and adjacent genes was subcloned in to pTA131 (*pyrE2*) to generate pTA252. The 1 kb *Bsm*I/*Xba*I fragment containing the intergenic region necessary for ARS activity was replaced by a *trpA* selectable marker for tryptophan biosynthesis to generate pTA266 (Allers *et al.,* 2004). This Δ*ori-pHV1/4*::*trpA* deletion construct was used to transform the H53 (Δ*pyrE2*, Δ*trpA*) strain to prototrophy for uracil and tryptophan. Integration at the *orc10* locus was verified by Southern blot (Figure 3.3, panel B, H220). Selection was relieved for uracil to encourage intramolecular recombination and pop-out. Counter-selection with 5-fluoroorotic acid (5-FOA) was then used to ensure loss of integrated pTA266. Selection for tryptophan was maintained to directly select for recombination events resulting in the deletion mutant. It was possible to delete *ori-pHV1/4* and by plating putative deletion strains on Hv-Ca+5-FOA +/- trp, the frequency of deletion events could be measured. Deletion events ($trpA^+$, 5-FOA-resistant cells) were outnumbered >1000-fold by events leading to restoration of the wild-type ($trpA^-$, 5-FOA-resistant cells). This indicates a strong bias for maintenance of the replication origin. The Δ*ori-pHV1/4*::*trpA* genotype was verified by a Southern blot of a genomic DNA diagnostic digest and by probing intact DNA displayed on a pulsed-field gel (PFG) with the deleted region (Figure 3.3, panel D, H230).

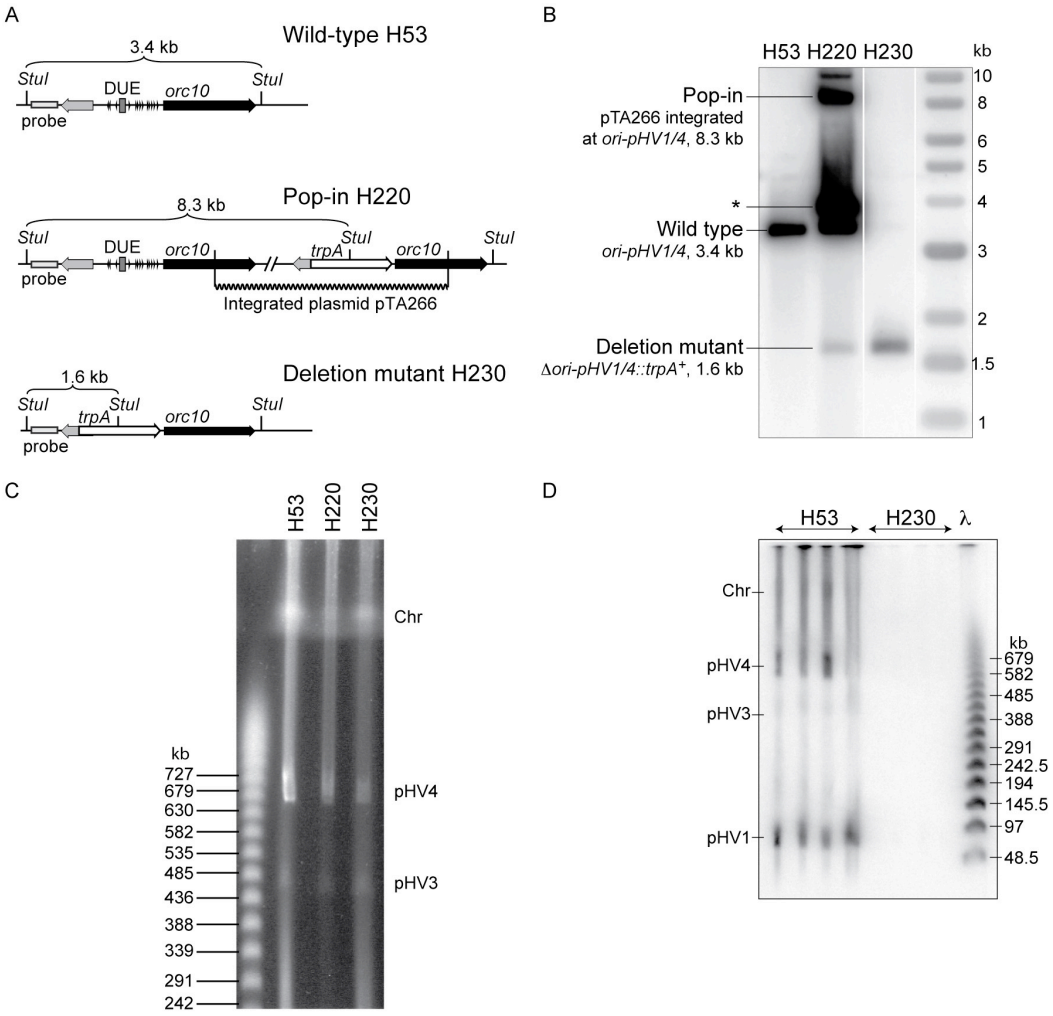


Figure 3.3: *ori-pHV1/4* deletion

(A) Deletion strategy for *ori-pHV1/4* showing the fragment sizes for diagnostic digests at each stage of strain construction.

(B) *Stu*I digest of genomic DNA from strains H53 (Δ*pyrE2*, Δ*trpA*), H220 (*ori-pHV1/4*+::[Δ*ori-pHV1/4*::*trpA*$^+$ *pyrE2*$^+$]) and H230 (Δ*ori-pHV1/4*::*trpA*$^+$, Δ*pyrE2*, Δ*trpA*), probed with DNA flanking the intergenic region. * Represents episomal DNA carrying the pHV1/4 origin, resulting from excision of the integrated plasmid.

(C) Ethidium bromide staining of intact *H. volcanii* DNA from strains H53, H220 and H230 showing the presence of pHV4 when *ori-pHV1/4* is deleted.

(D) Southern blot of intact *H. volcanii* DNA from H53 (Δ*pyrE2*, Δ*trpA*) and H230 (Δ*ori-pHV1/4*::*trpA*$^+$, Δ*pyrE2*, Δ*trpA*) probed with the *ori-pHV1/4* locus (pTA252 *Xba*I/*Bsm*I).

H230 (Δ*ori-pHV1/4*::*trpA*$^+$, Δ*pyrE2*, Δ*trpA*) still contains pHV1 and pHV4 (Figure 3.3, panel C, pHV1 not shown). Therefore pHV1 and pHV4 must be using alternative replication origins. Since the *ori-pHV1/4* deletion strain H230 showed no obvious growth defects, these alternative origins presumably act as sites of efficient DNA replication initiation.

3.2.2 Genetic screen for origins in a Δ*ori-pHV1/4* background

The genetic screen for ARS elements was repeated using the Δ*ori-pHV1/4* deletion strain H230 as the source of genomic DNA. A library of partially *Aci*I digested H230 fragments of 3-5 kb was cloned in to the compatible *Cla*I site of pTA131 (*pyrE2*) and used to transform H49 (Δ*pyrE2*, Δ*radA*) to prototrophy for uracil. Plasmid DNA from nine transformants was sequenced after passage through *E. coli* and all the inserts localised to a single region of the main chromosome shown in Figure 3.5. This region contains two divergently transcribed genes, *orc1* and a GTP-binding protein, separated by a 1360 bp intergenic region featuring two A/T-rich DUEs and several direct repeats. The DUE proximal to the predicted GTP-binding protein ORF has an A/T content of 61% and the DUE proximal to *orc1* has an A/T content of 56% compared to 33% for the entire genome. These features are similar to those found at *ori-pHV1/4* but at this origin the nucleotide sequence of the direct repeats (Figure 3.5, panel B) is 89% identical to the ORB sequence (Robinson *et al.,* 2004). In addition to *orc1*, a number of other genes related to DNA replication and repair are found in the vicinity of this ARS (Figure 3.5, panel A) These include the DP1 exonuclease subunit of the archaeal D family DNA polymerase (Cann *et al.,* 1998), the Hef helicase/endonuclease (Komori *et al.,* 2002), a homologue of the bacterial UvrC nucleotide excision repair protein (Van Houten *et al.,* 2005), an NAD-dependent DNA ligase (Zhao *et al.,* 2006), and the Hel308 helicase (Guy and Bolt, 2005). This is reminiscent of the replication genes located near to the *P. abyssi* origin in Figure 1.13 (Myllykallio *et al.,* 2000). The ARS library insert from plasmid pTA313 was used to probe a Southern blot of intact *H. volcanii* DNA displayed on a PFG (Figure 3.5, panel C). One band corresponding in size to the main chromosome was observed (2.9 Mb) (Charlebois *et al.,* 1991) so this origin was named *oriC-1*. A 1.1 kb *Sau*3AI/*Hind*III fragment of pTA313, containing only the intergenic region was subcloned in to pTA131 to generate pTA441 which could successfully transform strain H49 (Δ*pyrE2*, Δ*radA*) to prototrophy for uracil. pTA441 (*oriC-1*) transformants establish colonies ~3 days earlier than pTA250 (*ori-pHV1/4*) transformants. The minimal origin was further delimited using PCR to amplify a 692 bp fragment of the intergenic region (Figure 3.5, panel A, pCN11). This fragment could maintain the plasmid in an ARS assay (performed by Cédric Norais).

Figure 3.4: Features of *oriC-1*

(A) Sequence features of *oriC-1* and the regions used in ARS assays. See Figure 3.1 for key.

(B) Sequence of repeats (numbered in A) within the intergenic region adjacent to *orc1* are above the line. Below the line are sequences of repeats found at other confirmed and presumed archaeal origins. The species and relevant *cdc6/orc1* genes are *H. marismortui cdc6-4* (*Hmar-1-2*), *Halobacterium* sp. NRC-1 *orc7* (*NRC1-1 -5 2*), *Natronomonas pharaonis cdc6-1* (*Npha-1-2*), *S. solfataricus cdc6-1* (*Sso-1-2*) and *P. abyssi cdc6* (*Pab-1-4*). Repeat orientation is indicated by arrows and conserved positions are shaded. Amongst halophilic archaea, repeats surrounding the primary DUE feature a longer consensus sequence called the Halo-ORB (boxed). This contains the core mini-ORB and "G-string" elements found in other archaea and a halophile-specific "G-string". Analysis by Cédric Norais.

(C) Southern blot of intact *H. volcanii* DNA from strain H53 probed with the *ori-C1* locus (pTA313 *Xho*I/*Stu*I).

(D) H112 (Δ*radA*, Δ*pyrE2*) transformed with 1 µg of plasmid pTA441 (*ori-pHV1/4*). 100 µl of a 100-fold dilution plated on Hv-Ca and incubated at 45°C for 15 days.

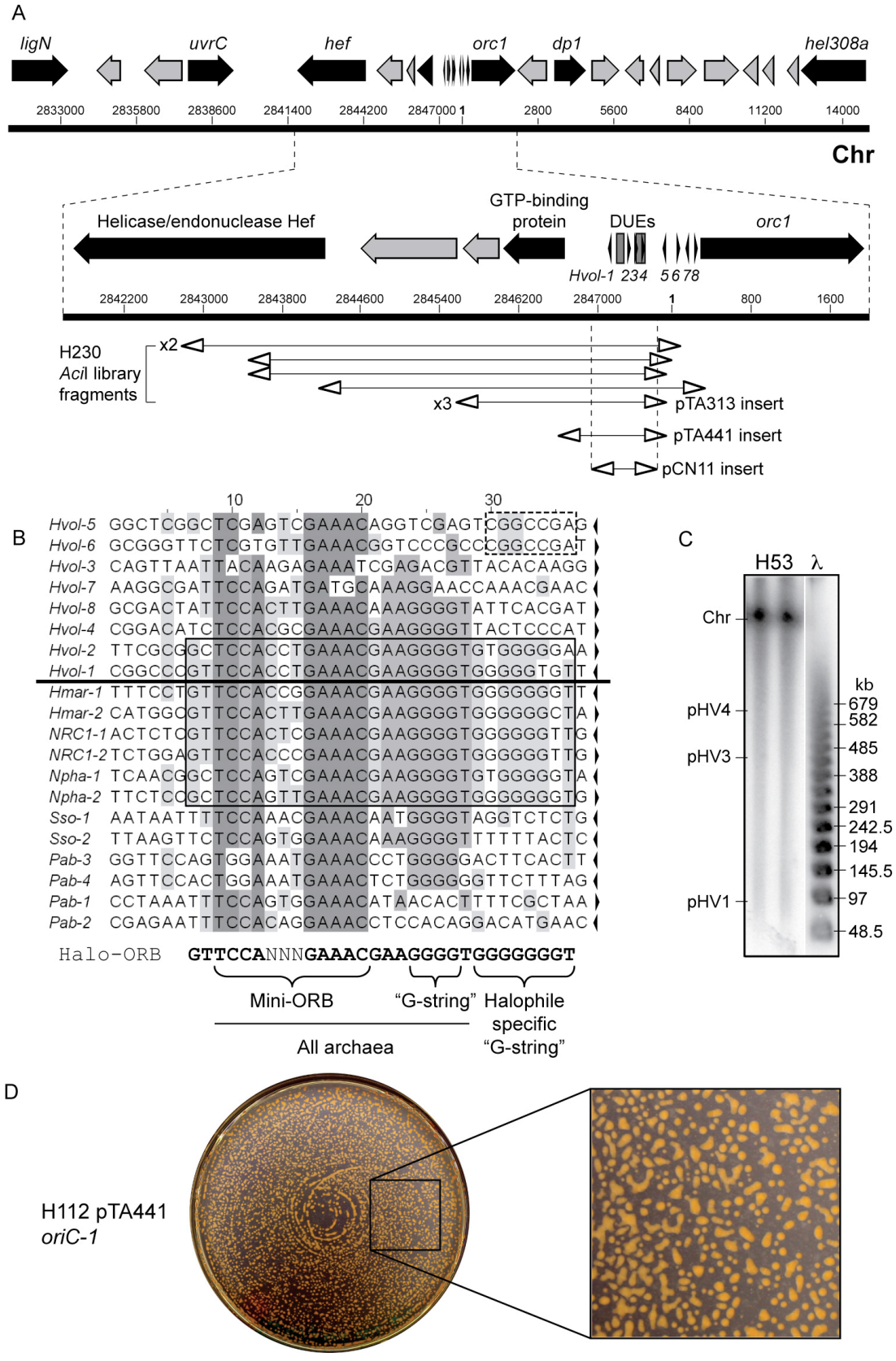


Figure 3.4: **Features of *oriC-1***

3.3 *In vivo* function of *ori-pHV1/4* and *oriC-1*

3.3.1 Two-dimensional agarose gel electrophoresis analysis

Two-dimensional agarose gel electrophoresis analysis (2-D agarose gel electrophoresis) was used to confirm that *ori-pHV1/4* and *oriC-1* act as replication origins *in vivo*. This method relies on identification of different replication intermediates by performing two perpendicular gel runs to separate DNA molecules by size and topology (see Figure 1.10).

H. volcanii is polyploid and there is an inverse correlation between cell density and chromosomal copy number (Breuert *et al.,* 2006). To maximise the potential number of replication bubbles, flow cytometry was used to measure the DNA content of *H. volcanii* cells in different growth phases.

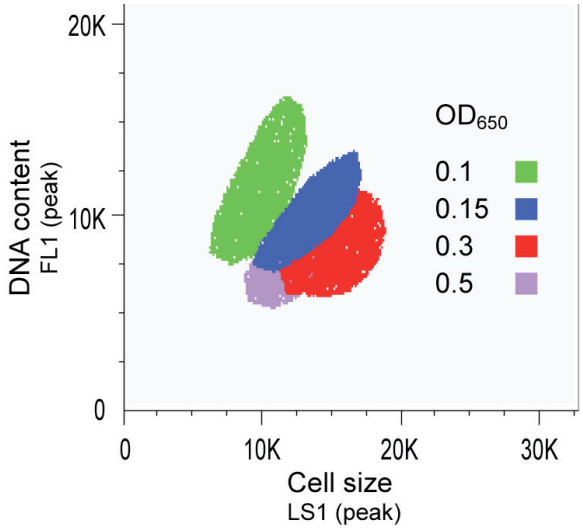


Figure 3.5: DNA content of *Haloferax volcanii* cells
Flow cytometry analysis of the DNA content of *H. volcanii* cells stained with acridine orange in different growth phases. The 510-580 nm FL1 parameter was chosen so that DNA-specific binding was measured as opposed to nucleic acid content. Data analysed using FlowJo.

Acridine orange binds DNA and RNA. This analysis showed that DNA content is highest when *H. volcanii* cells are in early exponential phase at an OD_{650} of 0.15. 40 ml of H678 (DS2) cells at an OD_{650} of 0.1 were used to prepare intact genomic DNA in agarose plugs. These plugs were digested with *Nru*I and 2-D agarose gel electrophoresis was carried out.

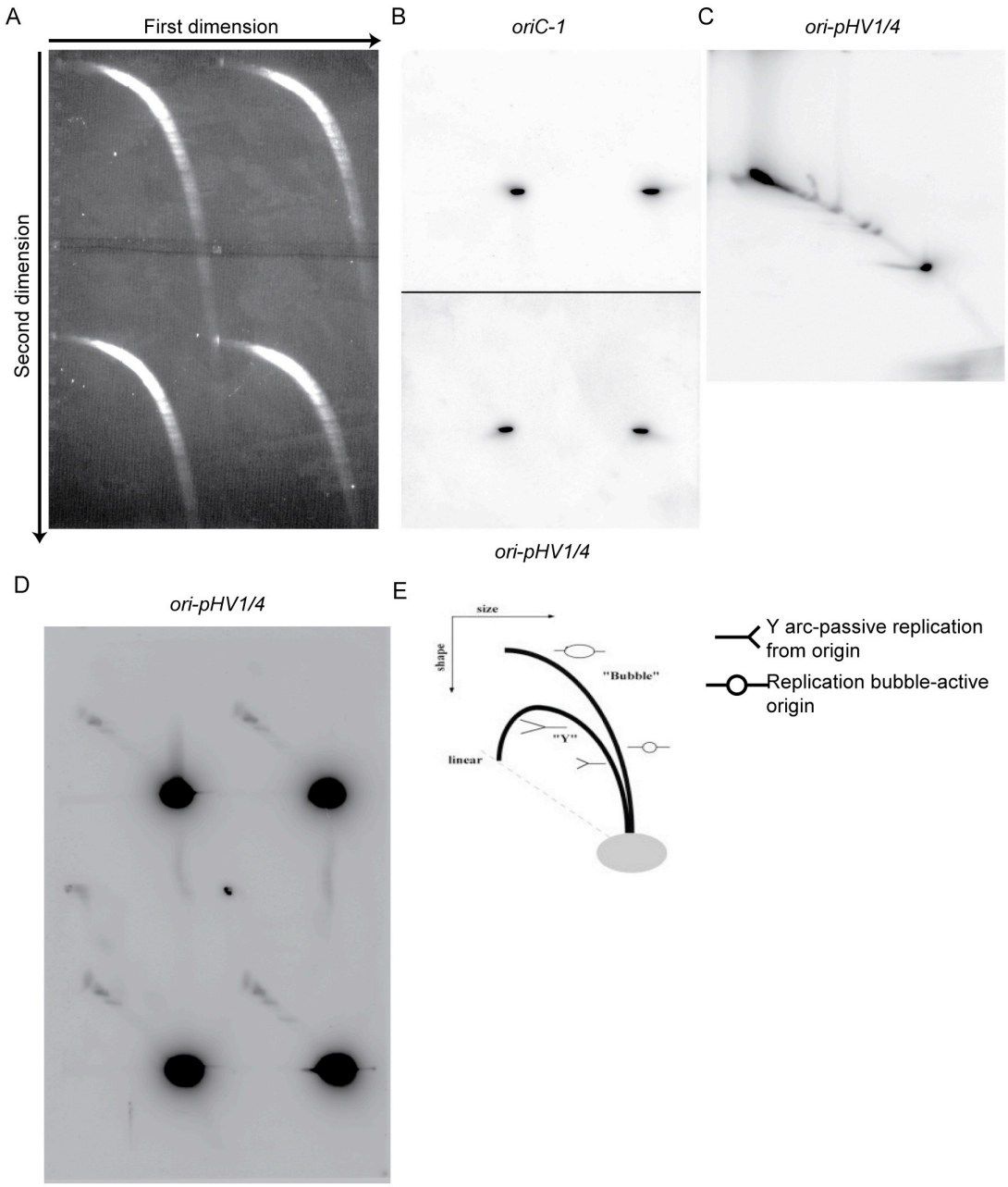


Figure 3.6: 2-D agarose gel electrophoresis analysis of *ori-pHV1/4* and *oriC-1*

(A) Ethidium bromide staining of *Nru*I digested H678 (DS2) genomic DNA.

(B) Southern blot of A, probed with *oriC-1* (pTA313 *Bam*HI/*Age*I) and *ori-pHV1/4* (pTA250 *Xho*I/*Eco*RV).

(C) Southern blot of *Ban*II digested H53 (Δ*pyrE2*, Δ*trpA*) genomic DNA after 2-D agarose gel electrophoresis performed by Cédric Norais. Probed with *ori-pHV1/4* (pTA250 *Xho*I/*Eco*RV).

(D) Southern blot of H9 (ΔpHV2) *Stu*I digested genomic DNA after 2-D agarose gel electrophoresis. Probed with *ori-pHV1/4* (p252 *Xba*I/*Bam*HI). One week phosphoimager exposure. Performed by Thorsten Allers.

(E) Diagram of replication intermediates after 2-D agarose gel electrophoresis.

Despite these sequences acting as replication origins in ARS assays, no replication fork bubbles were visible in this analysis. Many other 2-D agarose gel electrophoresis experiments were attempted with *H. volcanii* but none were successful in identifying origins of replication (T. Allers, unpublished observations). There has also been no success using this method with other halophiles, for example the closely related *Halobacterium sp. NRC-1* (H. Myllykallio, personal communication). 2-D agarose gel electrophoresis has been productive in other archaeal species so something specific to halophiles appears to prevent this method from working. A technical explanation could be that the high salt content of halophiles is crucial to maintaining replication intermediate structures. The cell lysis in this method might therefore be causing replication bubble degradation. Another possibility is that the polyploidy of halophiles (Breuert *et al.,* 2006) makes origin activity more difficult to detect in an asynchronous population. This could occur if the majority of origins across all the chromosome copies had already fired and obscured the minority of firing origins.

3.3.2 Replication initiation point mapping analysis

Since 2-D agarose gel electrophoresis was unsuccessful in *H. volcanii*, replication initiation point (RIP) mapping (see Figure 1.11) was used by Cédric Norais to confirm that *ori-pHV1/4* and *oriC-1* are active *in vivo*.

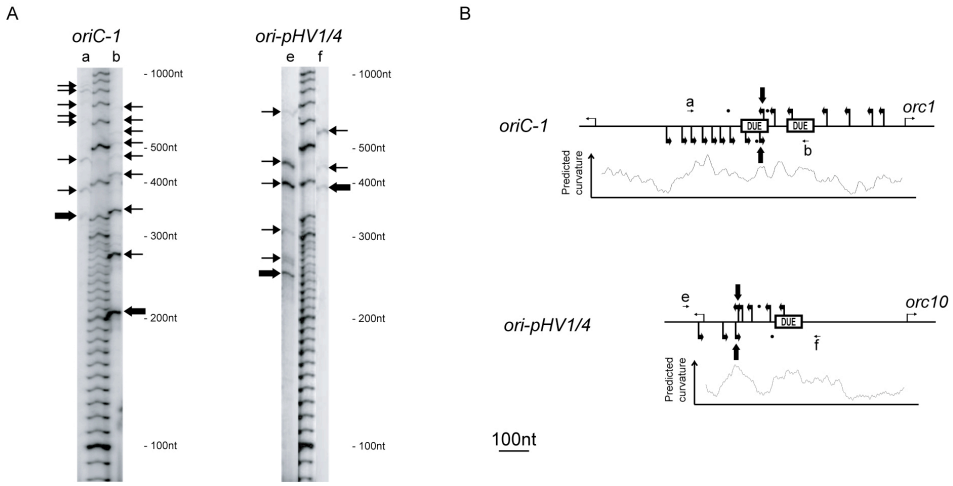


Figure 3.7: Mapping of initiation sites at *oriC-1* and *ori-pHV1/4*
(A) Primer extension reactions were performed with enriched replicating intermediate DNA fragments isolated by purifying single-stranded DNA on a BND-cellulose column. Extension reaction products were separated on a denaturing polyacrylamide gel and corresponding primers are indicated at the top (see table 2.3). Arrows refer to the 5′ end of amplification products and bold arrows indicate transition points. Performed by Cédric Norais.
(B) Position of primers and detected initiation sites on the origin sequence. The positions of duplex unwinding elements (DUE), the associated *orc* gene and the upstream ORF on the left are shown for both origins. Dots indicate weak amplification products obtained with control templates (linearised plasmid DNA, isolated from *E. coli* cells). Controls were also performed by Cédric Norais, data not available.

Clear transition points between the leading and lagging strand of replication are evident for *oriC-1* and *ori-pHV1/4*. For *oriC-1* the transition point is within the DUE while for *ori-pHV1/4* it is slightly upstream of the DUE motif. Both transition points co-localise with regions where DNA is predicted to be bent by the BEND.IT program which predicts DNA bending based on DNase I digestion of trinucleotides (Brukner *et al.,* 1995). These results confirm that *oriC-1* and *ori-pHV1/4* are functional origins of replication *in vivo*.

3.4 Characterisation of *ori-pHV1/4* and *oriC-1* ARS plasmids

To explain why only one origin was recovered from the initial genetic screen, the properties of these origins on episomes were explored. In the genetic screens for ARS, *H. volcanii* was transformed with a library of plasmids containing a range of sequences. It is likely that multiple plasmids enter any individual cell but only origin clones will produce transformants. The apparent dominance of the *ori-pHV1/4* sequence compared to *oriC-1* was investigated using co-transformation experiments. Co-transformations recapitulate aspects of the genetic screen using a simpler mixed population of two plasmids to transform *H. volcanii*. There are several reasons why one plasmid might transform more successfully than another in a co-transformation. A particular plasmid might be inherently more stable than another and would therefore be over-represented in the final population. Alternatively a particular plasmid could enable faster growth of transformants by replicating faster due to a more efficient origin. It is also possible that one plasmid could be dominant to others because of properties related to the cloned origin sequence.

H112 (Δ*pyrE2*, Δ*radA*) was transformed with a mixture of the *oriC-1* and *ori-pHV1/4* ARS plasmids isolated from the first and second genetic screens (0.5 µg each of pTA194 and pTA313). Co-transformants were restreaked and analysed by Southern hybridisation with a *pyrE2* probe for the presence of each ARS plasmid.

The vast majority of *oriC-1* and *ori-pHV1/4* co-transformants (21/23) contained the *ori-pHV1/4* plasmid pTA194, in most cases (17/23) this was the sole plasmid detectable. Only six transformants contained the *oriC1* plasmid pTA313 and in most cases (4/6) pTA194 was also present. To determine the fate of these ARS plasmids they were propagated by restreaking, both with maintenance of selection for the *pyrE2* marker and without selection. In all three cases analysed where both plasmids were present (Figure 3.8, panel B, transformants 3, 5, 10) only pTA194 remained after further propagation. This was not due to greater stability of pTA194 relative to pTA313, since transformants propagated without selection showed a marked loss of pTA194 (Figure 3.8, panel C). In contrast, pTA313 was well maintained in the transformant where it was present exclusively once selection was removed (Figure 3.8, panel C, transformant 8).

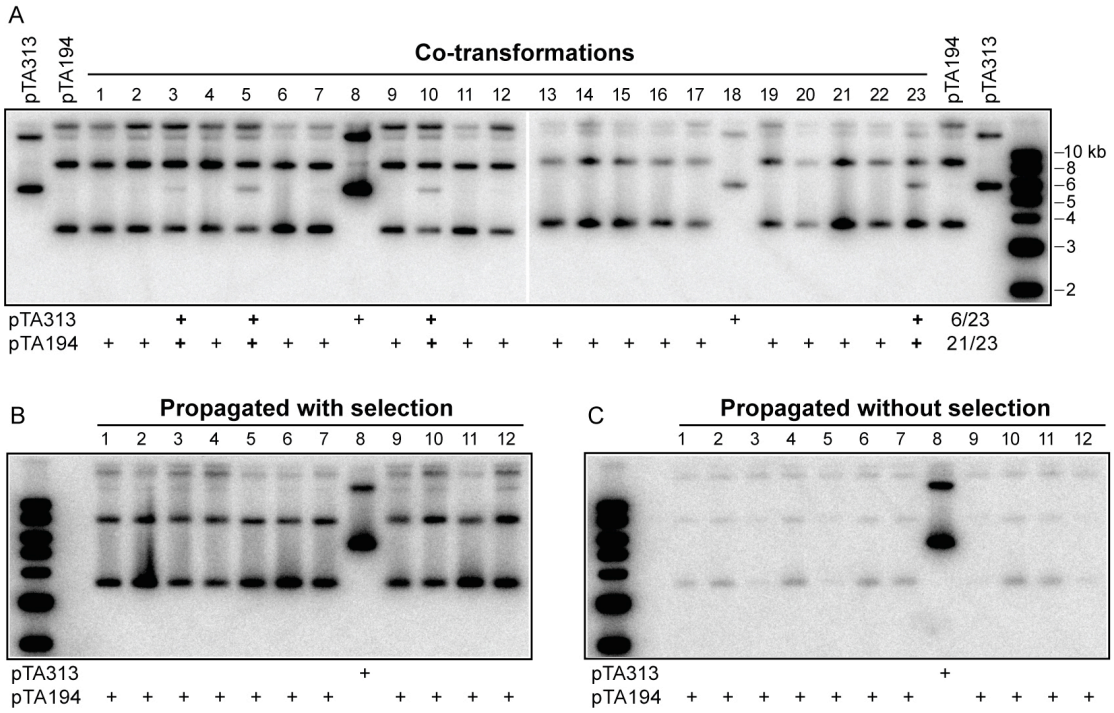


Figure 3.8: Co-transformation of *oriC-1* and *ori-pHV1/4* library clones
(A) H112 (Δ*pyrE2*, Δ*radA*) was transformed with 0.5 µg each of *oriC-1* (pTA194) and *ori-pHV1/4* (pTA313). Crude genomic DNA was isolated from 23 transformants grown in selective Hv-Ca broth. The Southern blot of this DNA was probed with *pyrE2* (pTA52 *Nco*I/*Xba*I) to locate all plasmid sequences.
(B) Transformants 1–12 were propagated by restreaking on selective Hv-Ca agar. DNA was isolated and probed as described for A.
(C) Transformants 1–12 were restreaked on non-selective Hv-YPC agar. DNA was isolated from Hv-YPC broth cultures of these transformants and probed as described for A.

These results match the outcome of the genetic screens and indicate that the *ori-pHV1/4* plasmid pTA194 is dominant to the *oriC* plasmid pTA313. The term "dominance" is used here to mean that a specific plasmid is over-represented in the transformant population.

In the initial genetic screen for ARS elements every sequenced transformant contained the *orc10* gene (see Figure 3.1), in contrast none of the sequenced transformants from the second screen that identified *oriC-1* contained *orc1* (see Figure 3.4). This could be an artifact of the screen resulting from restriction enzyme site preference or the size range chosen. In the initial co-transformation experiment the plasmids tested were pTA194 which was isolated from the first genetic screen and contains *orc10*, and pTA313 which was isolated from the second genetic screen

and does not contain *orc1*. Since stability or transformant colony establishment time can not explain the dominance of pTA194 over pTA313, the hypothesis that *orc10* was responsible was tested. If *orc10* was a negative regulator of origin firing at *oriC-1* it would explain the outcome of the genetic screens and the co-transformations shown in Figure 3.8. To test whether *orc10* was responsible for *ori-pHV1/4* plasmid dominance, co-transformation experiments were performed using all combinations of the intergenic origin region clones and the larger plasmids derived from the ARS genetic screens.

One factor that could influence co-transformation experiments is the size of the transforming DNA. The role of insert size in ARS assays was tested by transforming H112 with plasmids containing *oriC-1* clones of different sizes.

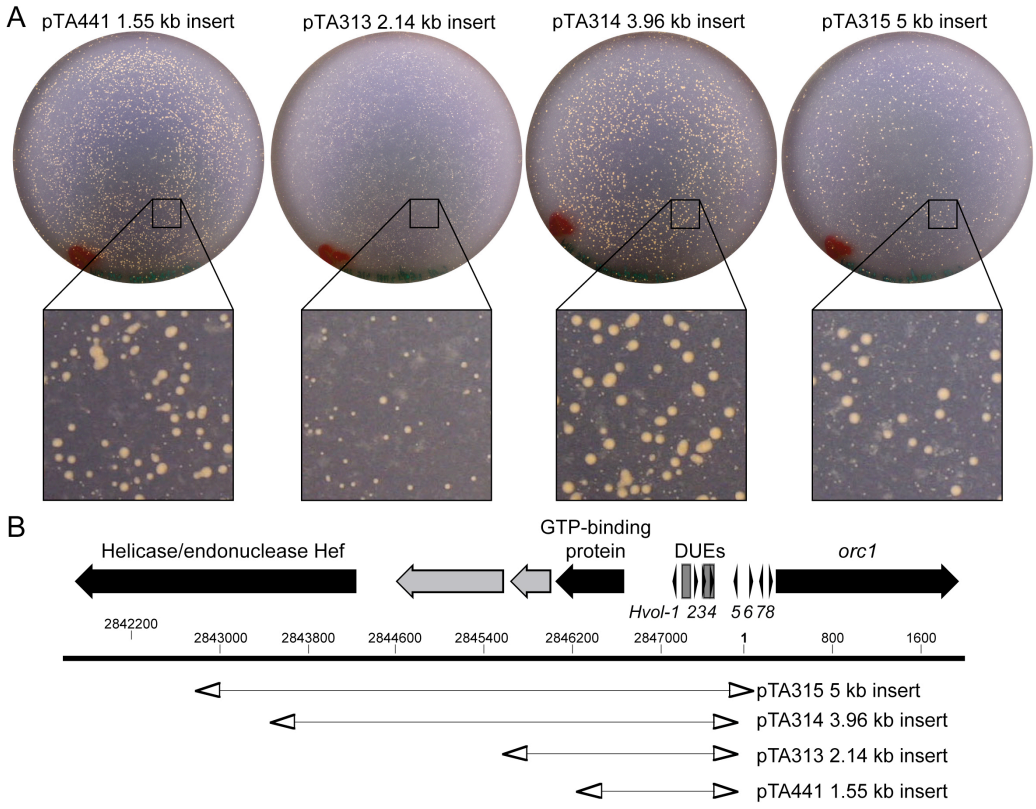


Figure 3.9: Effect of insert size on transformation efficiency
(A) H112 (Δ*radA*, Δ*pyrE2*) transformed with 1 µg of *oriC-1* based plasmids. 100 µl of a 1000-fold dilution plated on Hv-Ca and incubated at 45°C for 15 days.
(B) Sequences of different *oriC-1* clones used in A.

As the insert size increases the plasmid includes more of the flanking regions of *oriC-1*. Therefore the episomal *oriC-1* will be operating in a *cis* environment more like its chromosomal context. Figure 3.9 shows that in general, there is no significant

difference in transformation efficiency when *oriC-1* insert size is varied. pTA313 transformants produce smaller colonies but this is unlikely to be related to insert clone size since 1.55 and 5 kb *oriC-1* clones produce transformants with similar colony sizes. This means that a quantitative ARS assay for differentially sized origin clones is reasonable. At the extremes of insert size (pTA441, 1.1 kb and pTA315, 5 kb) there a difference in transformation efficiency. This has relevance for co-transformation experiments because the intergenic plasmids will have an advantage when co-transformed with the larger library insert plasmids. However, as shown by the dominance of the 8.6 kb *ori-pHV1/4* clone (pTA194) over the smaller 5.8 kb *oriC-1* clone (pTA313), this size advantage can be obscured by "dominant" plasmids.

Table 3.2: Plasmids used in co-transformation experiments

Plasmid	Origin	Clone type	Plasmid size (kb)
pTA441	*oriC-1*	Intergenic	5.1
pTA313	*oriC-1*	Library	5.8
pTA194	*ori-pHV1/4*	Library	8.6
pTA250	*ori-pHV1/4*	Intergenic	4.7

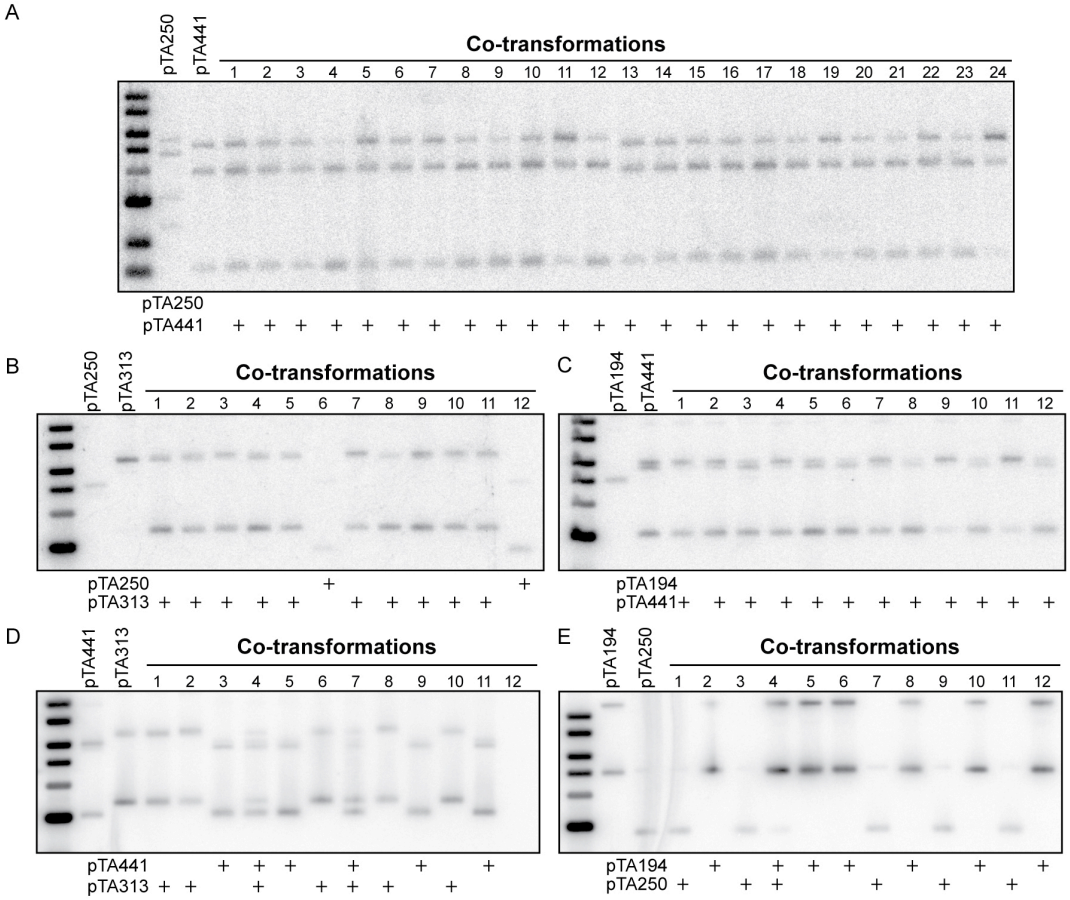


Figure 3.10: Co-transformations of *oriC-1* and *ori-pHV1/4* plasmids using intergenic and library clones

H112 (Δ*radA*, Δ*pyrE2*) transformed with 0.5 µg of each plasmid or single plasmid controls. For all panels the probe used was *pyrE2* (pTA52 *Nco*I/*Xba*I).

(A) Southern blot of *ori-pHV1/4*, (pTA250) and *oriC-1* (pTA441) co-transformants. Crude genomic DNA was extracted and digested with *Eco*RI/*Bsp*MI.

Panels B-E show undigested crude genomic DNA.

(B) Southern blot of *ori-pHV1/4* (pTA250) and *oriC-1* (pTA313) co-transformants.

(C) Southern blot of *ori-pHV1/4* (pTA194) and *oriC-1* (pTA441) co-transformants.

(D) Southern blot of *oriC-1*, (pTA441 and pTA313) co-transformants.

(E) Southern blot of *ori-pHV1/4*, (pTA194 and pTA250) co-transformants.

The dominance of these plasmids varies depending upon the size of the insert used. Panel A shows that the intergenic *oriC-1* plasmid (pTA441) is the only plasmid recovered from *oriC-1* (pTA441), *ori-pHV1/4* (pTA250) co-transformants (24/24). Therefore the intergenic *oriC-1* plasmid is dominant to the intergenic *ori-pHV1/4* plasmid. Panel B shows that the larger *oriC-1* library clone (pTA313) is also dominant to the intergenic *ori-pHV1/4* plasmid (pTA250) as it is the only plasmid present in 10/12 co-transformants. Panel C shows that the intergenic *oriC-1* plasmid (pTA441) is dominant to the library clone of *ori-pHV1/4* (pTA194) as it is present in all of the examined co-transformants (12/12). For these examples of clearly dominant plasmids (Figure 3.10, panels A-C) no co-transformants were detected. Panel D shows that the intergenic region and library clone of *oriC-1* are roughly equivalent based on their frequency of recovery from co-transformants (4/11 and 5/11 respectively). Panel E shows that the intergenic region (pTA250) and library clone of *ori-pHV1/4* (pTA194) are also roughly equivalent with individual recovery from co-transformants of 5/12 and 6/12 respectively. Co-transformants were recovered when the library clone and the intergenic region of the same origin were co-transformed for both *oriC-1* (Figure 3.10, panel D, lanes 4 and 7) and *ori-pHV1/4* (Figure 3.10, panel E, lane 4). The various dominance relationships can be summarised as follows:

Intergenic region		Library clone		Library clone		Intergenic region
oriC-1	>	*ori-pHV1/4*	>	*oriC-1*	>	*ori-pHV1/4*
pTA441		pTA194		pTA313		pTA250

Library clone		Intergenic region
oriC-1	≈	*oriC-1*
pTA313		pTA441

Library clone		Intergenic region
ori-pHV1/4	≈	*ori-pHV1/4*
pTA194		pTA250

Based on the behaviour of these *H. volcanii* origin based plasmids in co-transformations, the only sequence that might be expected to be isolated from the genetic screen over and above the dominant *ori-pHV1/4* clone pTA194, would be intergenic *oriC-1*. This sequence is 1.55 kb and so would not have been included in the 4-8 kb library used in the initial genetic screen. It is striking that when the intergenic origin sequences are used in a co-transformation, the dominance relationship is inverse to that seen from the longer sequences isolated from the libraries. This supports the hypothesis that *orc10* inhibits replication at other origins.

3.4.1 Stability of *ori-pHV1/4* and *oriC-1* ARS plasmids

The stability of the intergenic *oriC-1* and *ori-pHV1/4* ARS plasmids was tested in a quantitative manner. Cultures of H112 (Δ*radA*, Δ*pyrE2*) transformed with *oriC-1*, (pTA441) or *ori-pHV1/4* (pTA250) were propagated for ~25 generations in non-selective Hv-YPC broth. At regular intervals, the fraction of ura^{+} cells (indicative of ARS plasmid presence) was determined by plating on selective and non-selective media. As shown in Figure 3.11, the *ori-pHV1/4* plasmid pTA250 was significantly less stable than the *oriC-1* plasmid pTA441. Plasmid loss was calculated to be 16% per generation for pTA250 and 9% per generation for pTA441. This was reflected in qualitative stability tests performed by propagation on non-selective solid media (Hv-YPC) followed by restreaking on selective solid media (Hv-Ca) (Figure 3.11, panel B). Nine transformants for each ARS plasmid were tested using this method and both plasmids showed 100% maintenance. However, *ori-pHV1/4* transformant restreaks took longer than *oriC-1* transformants to present colonies of the same size. This difference was greater than the ~48 hours difference it took for the original transformants to present similar sized colonies.

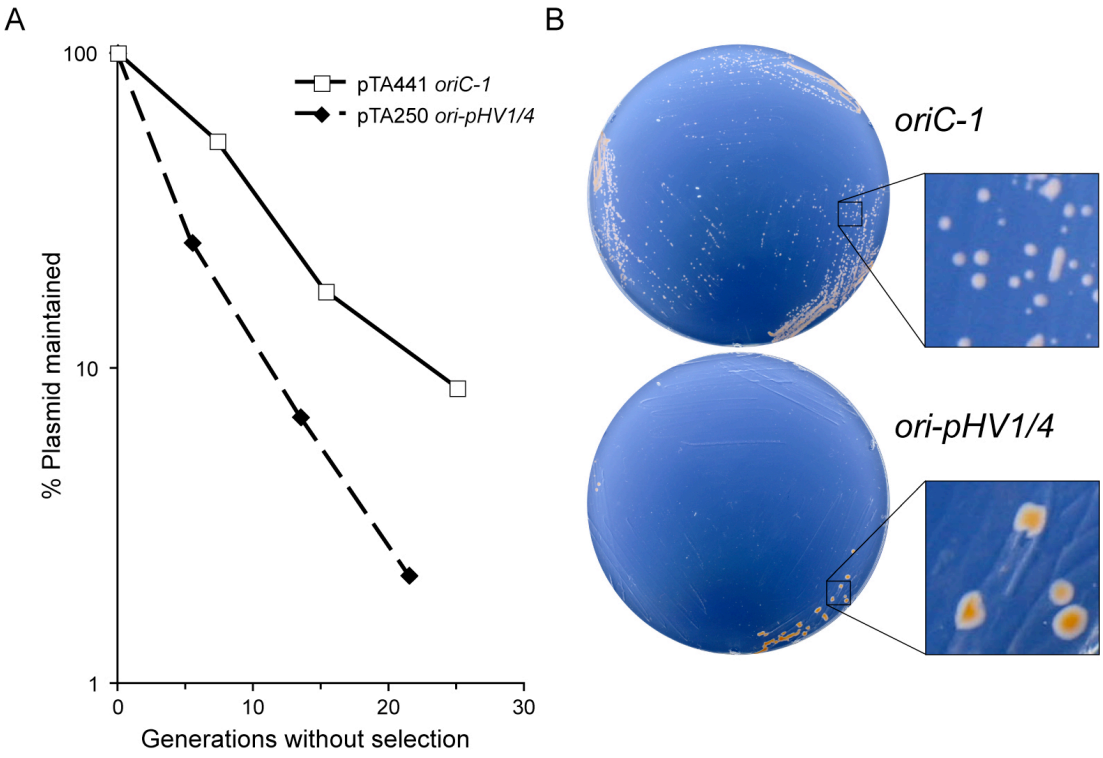


Figure 3.11: Stability of *oriC-1* and *ori-pHV1/4* plasmids

Figure 3.11: Stability of *oriC-1* and *ori-pHV1/4* plasmids

(A) H112 (Δ*radA*, Δ*pyrE2*) containing *ori-pHV1/4* (pTA250) or *oriC-1* (pTA441) was grown in non-selective Hv-YPC broth. At regular intervals aliquots were plated on selective (Hv-Ca) and non-selective (Hv-YPC) agar to determine the fraction of ura^+ cells.
(B) H112 (Δ*radA*, Δ*pyrE2*) containing *ori-pHV1/4* (pTA250) or *oriC-1* (pTA441) restreaked on selective solid Hv-Ca media following propagation on non-selective solid Hv-YPC media. Plate pictures were taken after 8 days of growth at 45°C for *oriC-1* transformants and 15 days for *ori-pHV1/4* transformants.

3.5 Origin identification using bioinformatics

During this work the preliminary genome sequence for *H. volcanii* became available from TIGR. This enabled nucleotide disparity analysis of the *H. volcanii* replicons. This method presents *in silico* predictions of putative origins based on strand bias and has been used successfully to identify origins of replication in other archaeal species (see 1.5).

The nucleotide disparity of the *H. volcanii* genome sequence was analysed using the Z-CURVE algorithm by Cédric Norais (http://tubic.tju.edu.cn/zcurve/) (Zhang and Zhang, 1994; Zhang and Zhang, 2005). Z-curves are three-dimensional representations of DNA with each axis/component representing a biological property of the sequence. The x_n component displays the purine/pyrimidine distribution of bases (A or G/C or T) along the sequence; the y_n component displays the distribution of amino/keto (A or C/G or T) base types and the z_n component displays the distribution of bases with weak or strong hydrogen bonding (A or T/G or C). Any DNA sequence can be described using the three components as coordinates and as with other nucleotide disparity algorithms the inflection points mark putative replication origin sites. Z-curve components can be combined to construct nucleotide disparity skews. For example, x_n and y_n combine to produce a GC disparity skew. Two-dimensional representations of nucleotide disparity skews for *H. volcanii* replicons are shown in Figure 3.12.

The nucleotide disparity curves show inflection points at the two previously identified origins. The G/C-skew of the main chromosome has two additional peaks, one associated with the *orc11* and *orc14* genes and one with *orc5*.

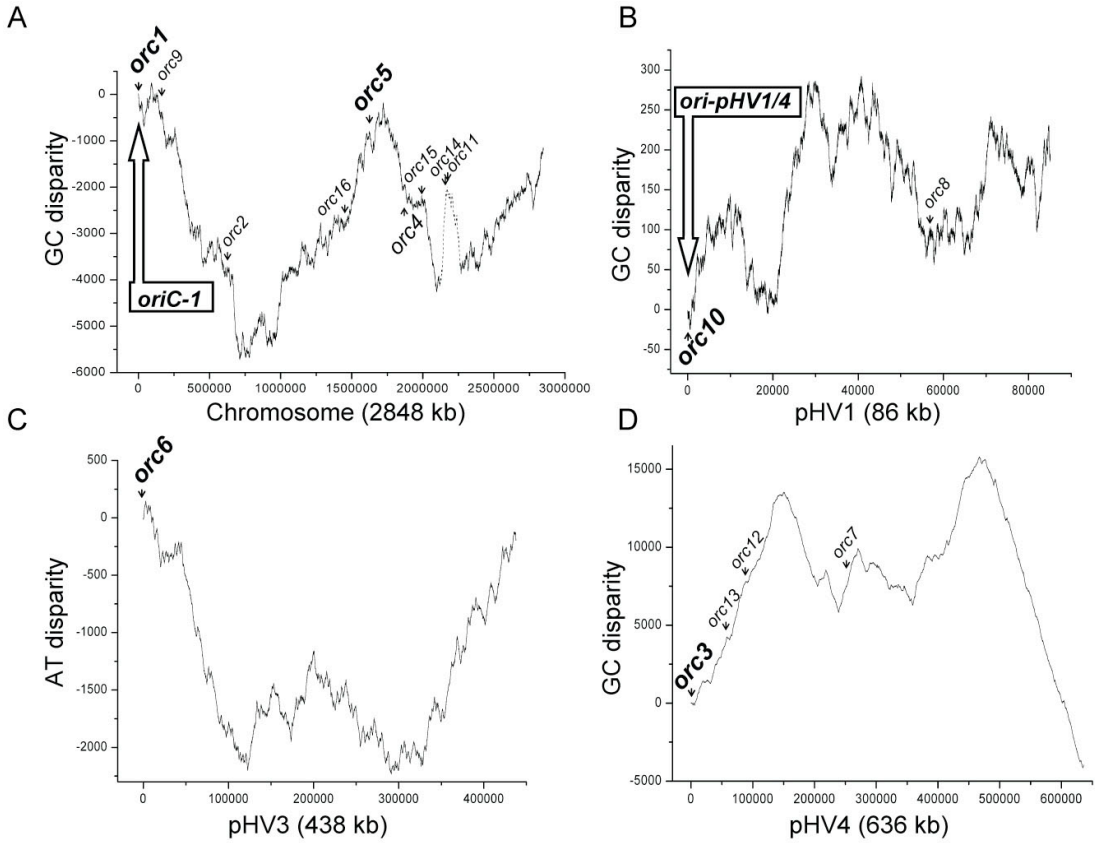


Figure 3.12: Nucleotide disparity curves for *Haloferax volcanii* replicons
(A) Main chromosome G/C skew showing the *orc* gene positions, *orc* genes close to a skew inflection point are bolded. The previously identified origin *oriC-1* is marked. A putative prophage sequence with high A/T content results in the sharp peak of nucleotide disparity indicated by the dotted line (see text for details).
(B) pHV1 G/C skew showing the *orc* gene positions, the *orc* gene close to the skew inflection point is bolded and the previously identified origin *ori-pHV1/4* is marked.
(C) and (D) pHV3 and pHV4 nucleotide skews respectively, *orc* gene positions are shown and *orc* genes close to skew inflection points are bolded. Z curve analysis by Cédric Norais.

The peak associated with *orc11*/*14* is within a sharply delineated chromosome region that has a lower G/C content (50%) than the genome average of 66%. This region is a proposed prophage because it contains putative viral genes such as homologues of the XerB/C site-specific recombinases (Hartman *et al.,* in preparation). The *orc5*-associated peak suggests the presence of a second replication origin on the main chromosome which was named *oriC-2*. pHV4 and pHV3 also show inflection points associated with *orc* genes. These three putative replication origins were subsequently cloned and tested for origin activity.

3.5.1 *oriC-2*

The region containing the *orc5*-associated skew inflection point was cloned to test the prediction that it was a second chromosomal replication origin. A 9.25 kb *Sac*I fragment from a H9 genomic DNA library was cloned in pBluescript II SK+ at the *Sac*I site to generate pTA416 by Zhenhong Duan (University of Nottingham, Nottingham). This fragment contains two divergently transcribed genes, *orc5* and a putative *RAD25/XPB*-related helicase (see Chapter 5), separated by a 2094 bp intergenic region featuring an A/T-rich DUE (55% A/T vs. 33% for the entire genome) and several direct repeats with similarity to the ORB consensus sequence (Figure 3.13, panel B). These features are similar to those found at *oriC-1* and *ori-pHV1/4*. This region is 6 kb from the *rrnB* operon which is orientated away from the putative replication origin.

A 2.4 kb fragment containing the intergenic region was amplified by PCR (primers F-1cdc and R-1radh) and cloned in the blunt *Eco*RV site of pTA131 (*pyrE2*) to generate pTA612. This intergenic *oriC-2* clone was used to transform H112 (Δ*pyrE2*, Δ*radA*) to test for ARS activity. A high frequency of transformants was observed which indicates that this region supports ARS activity. After 15 days of incubation at 45°C the transformant colonies were significantly smaller than those seen with H112 transformed with *ori-pHV1/4* (pTA250) or *oriC-1* (pTA441) clones. A Southern blot of intact *H. volcanii* DNA displayed on a PFG confirmed that this origin of replication is located on the main chromosome (Figure 3.13, panel C).

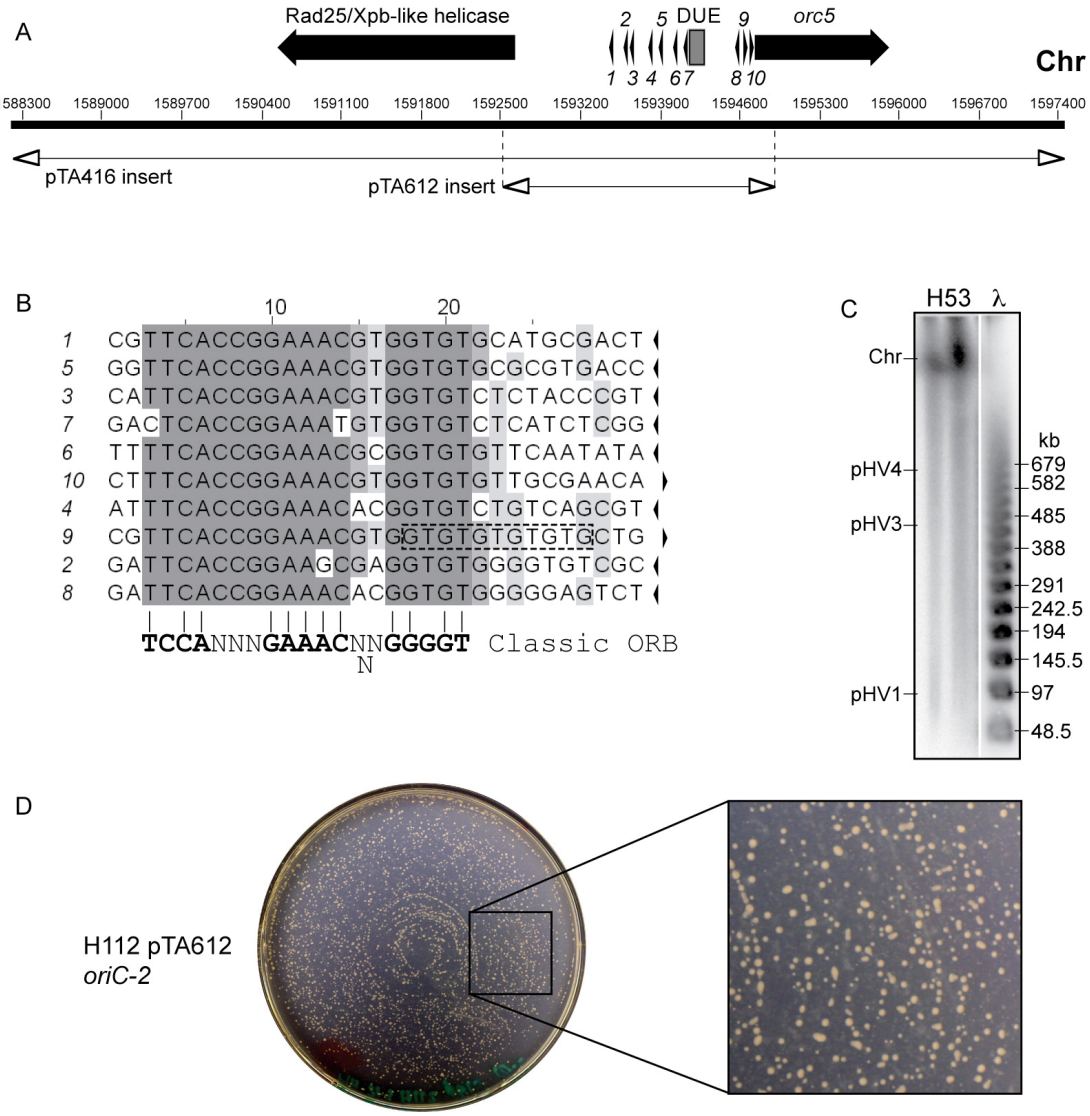


Figure 3.13: Features of *oriC-2*

(A) Sequence features of *oriC-2* and the regions used in ARS assays. See Figure 3.1 for key.

(B) Sequence of repeats (numbered in A) within the intergenic region adjacent to *orc5*. Repeat orientation is indicated by arrows and conserved positions are shaded. Analysis by Cédric Norais.

(C) Southern blot of intact *H. volcanii* DNA from strain H53, probed with the *oriC-2* locus (pTA419 *Bam*HI/*Eco*RI).

(D) H112 (Δ*radA*, Δ*pyrE2*) transformed with 1 µg of plasmid pTA612 (*oriC-2*). 100 µl of a 100-fold dilution plated on Hv-Ca and incubated at 45°C for 15 days.

3.5.2 *ori-pHV3*

The skew inflection point associated with *orc6* on pHV3 was also tested for origin activity. The 793 bp intergenic sequence adjacent to *orc6* features two DUEs (71% A/T for the smaller DUE and 67% A/T for the larger DUE vs. 33% for the entire genome) and several direct repeats with similarity to the ORB consensus sequence. A 693 bp intergenic PCR fragment including the DUEs and repeats was cloned in pTA131 (*pyrE2*) to generate pCN26/pTA734 (by Cédric Norais). This clone was used to transform H112 (Δ*pyrE2*, Δ*radA*) to test for ARS activity. A high frequency of transformants was observed but as with *oriC-2* transformants, the colonies were small after 15 days of incubation. The pTA734 insert was used to probe a blot of *H. volcanii* DNA displayed on a PFG. This confirmed that this replication origin is located on the pHV3 replicon and it was named *ori-pHV3*.

3.5.3 *ori-pHV4-2*

The final putative replication origin identified based on nucleotide skew inflection point and *orc* gene association was on pHV4. The 688 bp intergenic region adjacent to *orc3* on pHV4 features a DUE (69% A/T vs. 33% for the entire genome) and multiple direct repeats that are different from the ORB consensus sequence. These repeats are also unlike the repeats identified at *ori-pHV1/4*. Since *ori-pHV1/4* is located on pHV4 this putative origin was named *ori-pHV4-2*. This putative origin was tested for ARS activity by cloning a 592 bp PCR fragment of the intergenic sequence in pTA131 (*pyrE2*) to generate pCN27/pTA735 (cloning by Cédric Norais). Significantly fewer transformants were obtained when pTA735 (*ori-pHV4-2*) was used to transform H112 (Δ*pyrE2*, Δ*radA*) in the ARS assay compared with the other origins identified in this work. Plating 10% of undiluted transformants resulted in a maximum of 10 colonies whereas the other origins described produced confluent plates when undiluted. ARS assay transformant colonies were of a similar size to *ori-pHV3* and *oriC-2* transformants. A PFG probed with the *ori-pHV4-2* sequence showed that this origin is located on the 2.9 Mb main chromosome which suggests that our laboratory strains and/or the strain sequenced by TIGR have undergone genome rearrangements.

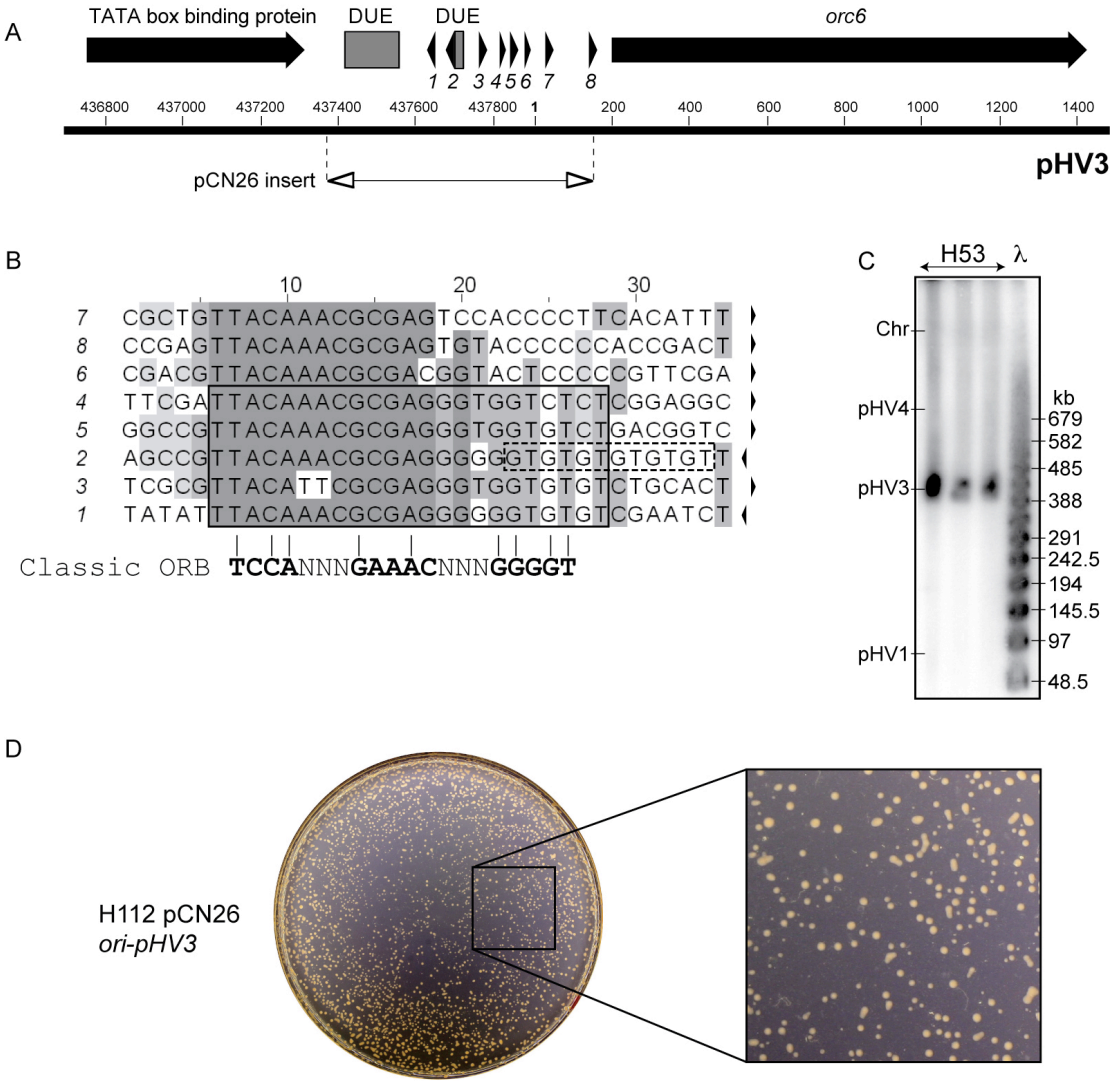


Figure 3.14: Features of *ori-pHV3*

(A) Sequence features of *ori-pHV3*. See Figure 3.1 for key.

(B) Sequence of repeats (numbered in A) within the intergenic region adjacent to *orc6*. Repeat orientation is indicated by arrows and conserved positions are shaded. Analysis by Cédric Norais.

(C) Southern blot of intact *H. volcanii* DNA from strain H53, probed with the *ori-pHV3* locus (pTA734 *Bam*HI/*Eco*RI).

(D) H112 (Δ*radA*, Δ*pyrE2*) transformed with 1 µg of plasmid pTA734 (*ori-pHV3*). 100 µl of a 100-fold dilution plated on Hv-Ca and incubated at 45°C for 15 days.

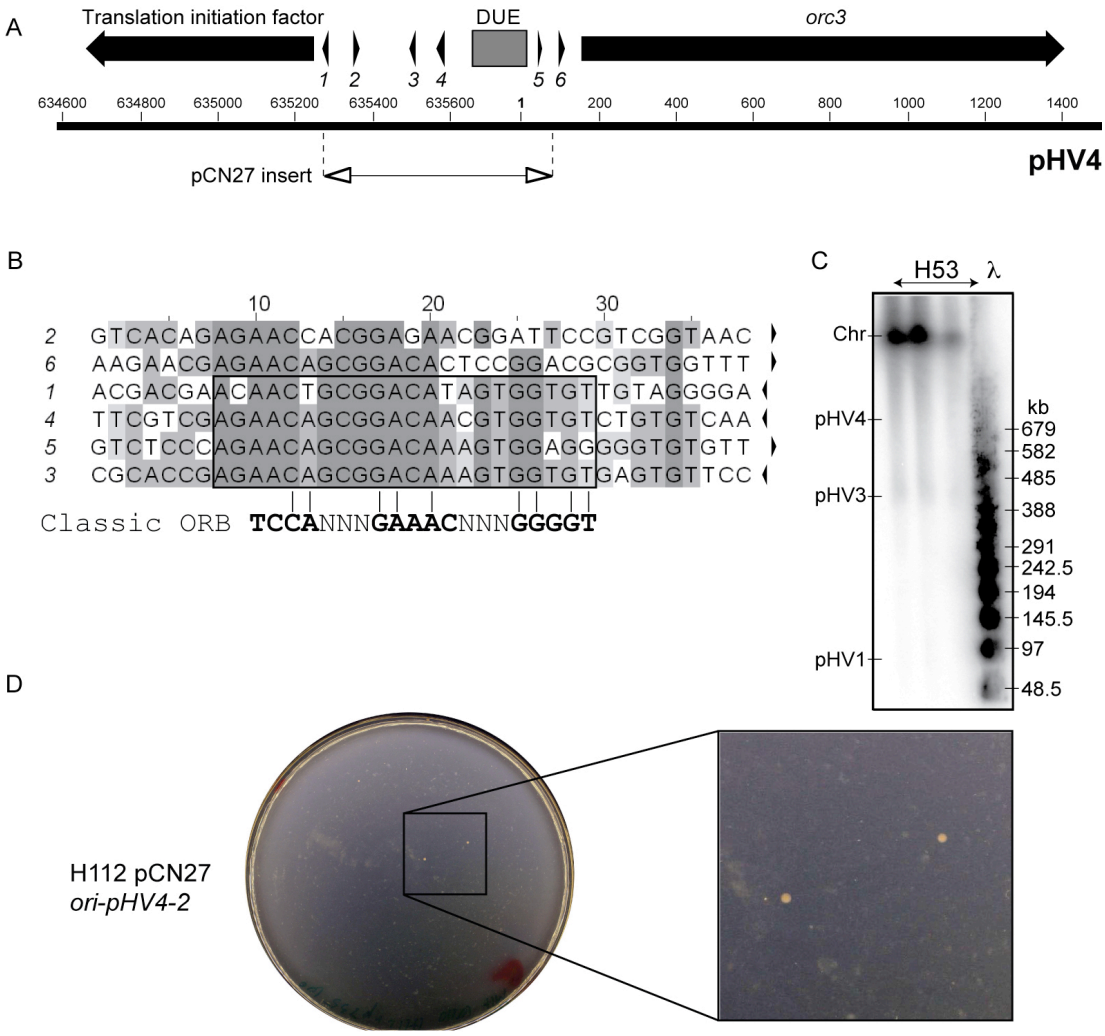


Figure 3.15: Features of *ori-pHV4-2*

(A) Sequence features of *ori-pHV4-2*. See Figure 3.1 for key.

(B) Sequence of repeats (numbered in A) within the intergenic region adjacent to *orc3*. Repeat orientation is indicated by arrows and conserved positions are shaded. Analysis by Cédric Norais.

(C) Southern blot of intact *H. volcanii* DNA from strain H53, probed with the *ori-pHV4-2* locus (pTA735 *Bam*HI/*Eco*RI).

(D) H112 (Δ*radA*, Δ*pyrE2*) transformed with 1 µg of plasmid pTA735 (*ori-pHV3*). 100 µl of undiluted cells plated on Hv-Ca and incubated at 45°C for 15 days.

3.6 Autonomously replicating sequence assay controls

The ARS assay was performed with the five intergenic origin clones numerous times and always gave similar results for transformant numbers and colony size. The transformation plates shown in panel D of Figures 3.1, 3.4, 3.13-3.15 were representative. As a negative control, the intergenic region adjacent to *orc4* was tested for origin activity (*oriC-X*). This region does not coincide with a nucleotide skew inflection point (Figure 3.12, panel A) and does not include the common features of *H. volcanii* origins (DUE, ORB sequences).

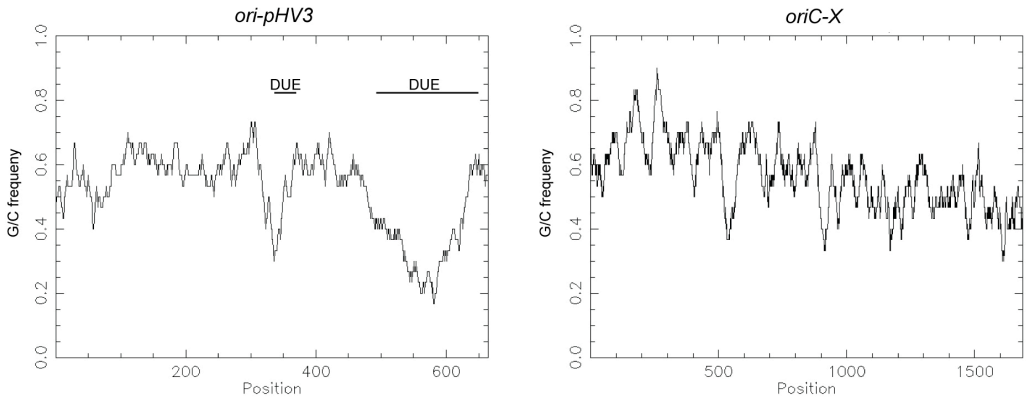


Figure 3.16: G/C frequency for *ori-pHV3* and *oriC-X*
G/C frequency plot for the intergenic region of pTA734 (*ori-pHV3*) and pTA611 (*oriC-X*) using the "freak" program (http://bioweb2.pasteur.fr). The two DUE regions of *ori-pHV3* are shown whereas no DUE is apparent at *oriC-X*.

A *Sac*I/*Nsp*I library was made from H1 and an 8.5 kb fragment containing the *orc4* locus was cloned in pTA176 (pUC19) at the *Sac*I/*Sph*I sites. The 4.7 kb *Xho*I/*Hin*dIII fragment of this plasmid was subcloned in pTA131 to generate pTA451 (cloning by Thorsten Allers). To obtain the intergenic region for ARS testing, the 1.7 kb *Apa*I/*Bgl*I intergenic fragment was subcloned in to pTA131 (*pyrE2*) to generate pTA611. No transformants were observed when *oriC-X* was used to transform H112 (Δ*pyrE2*, Δ*radA*) so not every *orc* gene in *H. volcanii* is associated with a DNA replication origin. Additional *H. volcanii* sequences that were available in the laboratory and not expected to function as origins of replication (*bgaHv*, *leuB* and the intergenic region adjacent to *polX*) were also cloned in pTA131 (*pyrE2*) to act as negative controls. No transformants were ever observed when these plasmids were used in ARS assays.

To verify that the *H. volcanii* origin clones were maintained as episomally replicating plasmids, total transformant DNA was analysed. Crude DNA was prepared from transformants grown in selective broth (Hv-Ca) and digested with *Eco*RI to linearise the transforming plasmid. The Southern blot of this digested DNA was probed with pBluescript II SK+ to identify all foreign DNA. Figure 3.17, panel A shows that all five transforming plasmids containing *H. volcanii* origins of replication linearised to the correct size. Therefore these plasmids replicate episomally in *H. volcanii*. The additional band (*) for the *oriC-2* transformant will be discussed in Chapter 4.

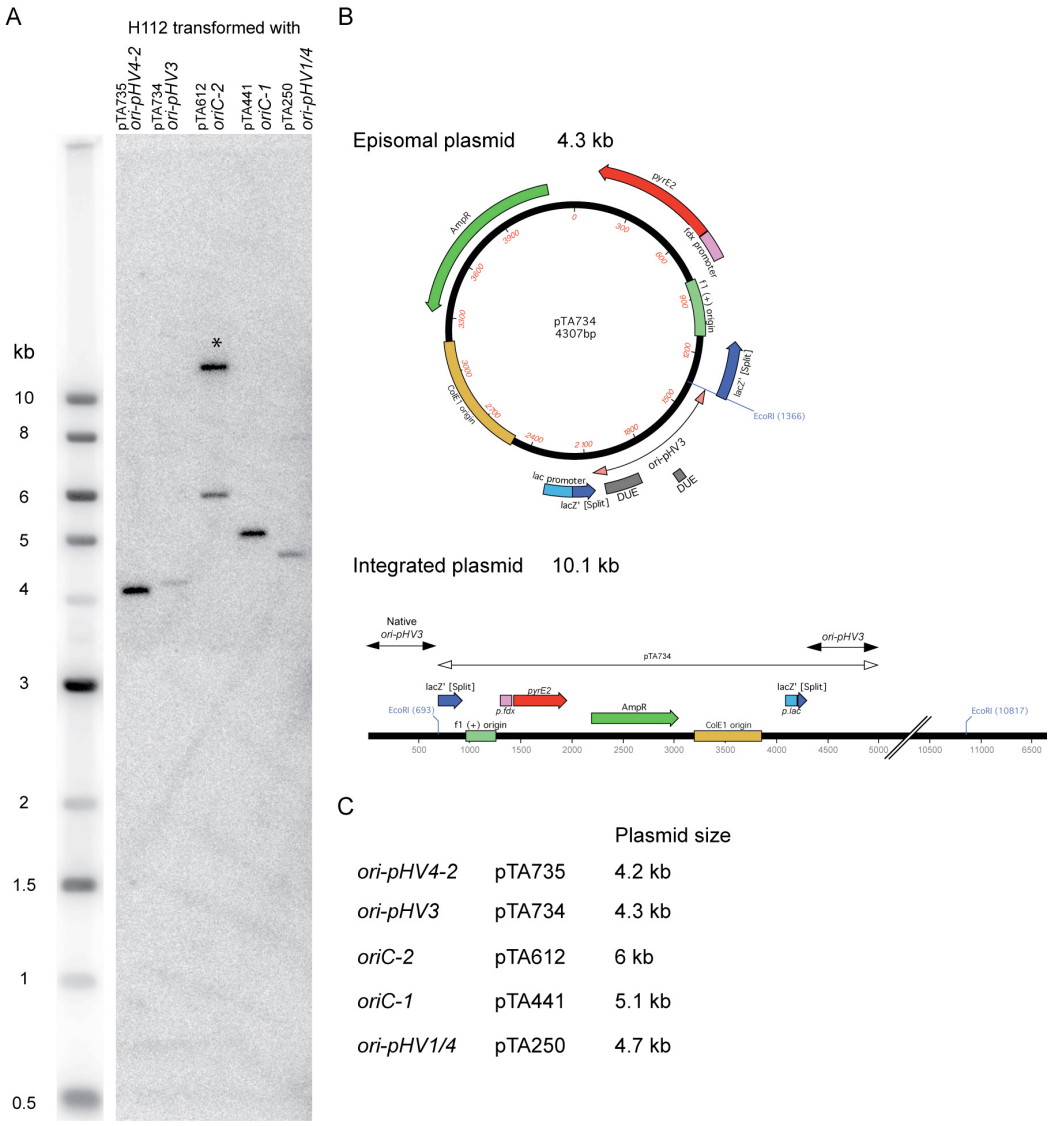


Figure 3.17: ARS assay controls

(A) Total DNA from H112 transformed with *H. volcanii* origin clones, digested with *Eco*RI and probed with pBluescript II SK+. * Will be discussed in Chapter 4.

(B) Example of expected fragments for episomal plasmid linearisation or integration for pTA734.

(C) Origin clone plasmid sizes.

3.6.1 *In vivo* confirmation of *oriC-2*, *ori-pHV3* and *ori-pHV4-2* function

As with *oriC-1* and *ori-pHV1/4*, RIP mapping was used to confirm that the three origins identified through bioinformatics functioned *in vivo*.

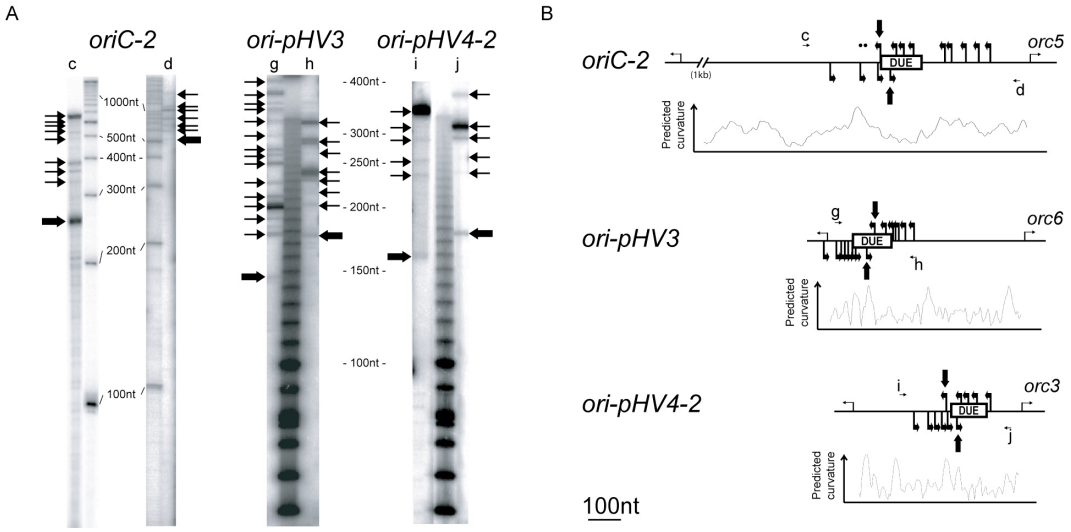


Figure 3.18: Mapping of initiation sites at *oriC-2*, *ori-pHV3* and *ori-pHV4-2*
(A) Primer extension reactions were performed with enriched replicating intermediate DNA fragments and separated on a denaturing polyacrylamide gel. Corresponding primers are indicated at the top of the gels, arrows refer to the 5′ end of amplification products and bold arrows indicate transition points. Performed by Cédric Norais.
(B) Primers and detected initiation sites along the sequence. The positions of duplex unwinding elements (DUE), the associated *orc* gene and the upstream ORF on the left are shown for each origin. Dots indicate weak amplification products obtained with control templates (linearised plasmid DNA, isolated from *E. coli* cells). Controls were also performed by Cédric Norais, data not available.

As shown by the clear transition points, these three origins are functional in their chromosomal context.

A combination of genetic and bioinformatic techniques resulted in the identification of five replication origins in *H. volcanii*. Each origin was functional in ARS assays although *ori-pHV4-2* ARS activity is greatly reduced compared to the four other origins. RIP mapping demonstrated that each *H. volcanii* replication origin is a functional origin *in vivo*. Importantly the circular main chromosome of *H. volcanii* was shown to contain multiple origins of replication. This is the first example of multiple origins on a single replicon in euryarchaea.

3.7 Comparative analysis of DNA replication origin ARS plasmids

3.7.1 Efficiency

As noted previously, the *H. volcanii* origin-based plasmids produce transformant colonies of different sizes in the ARS assay.

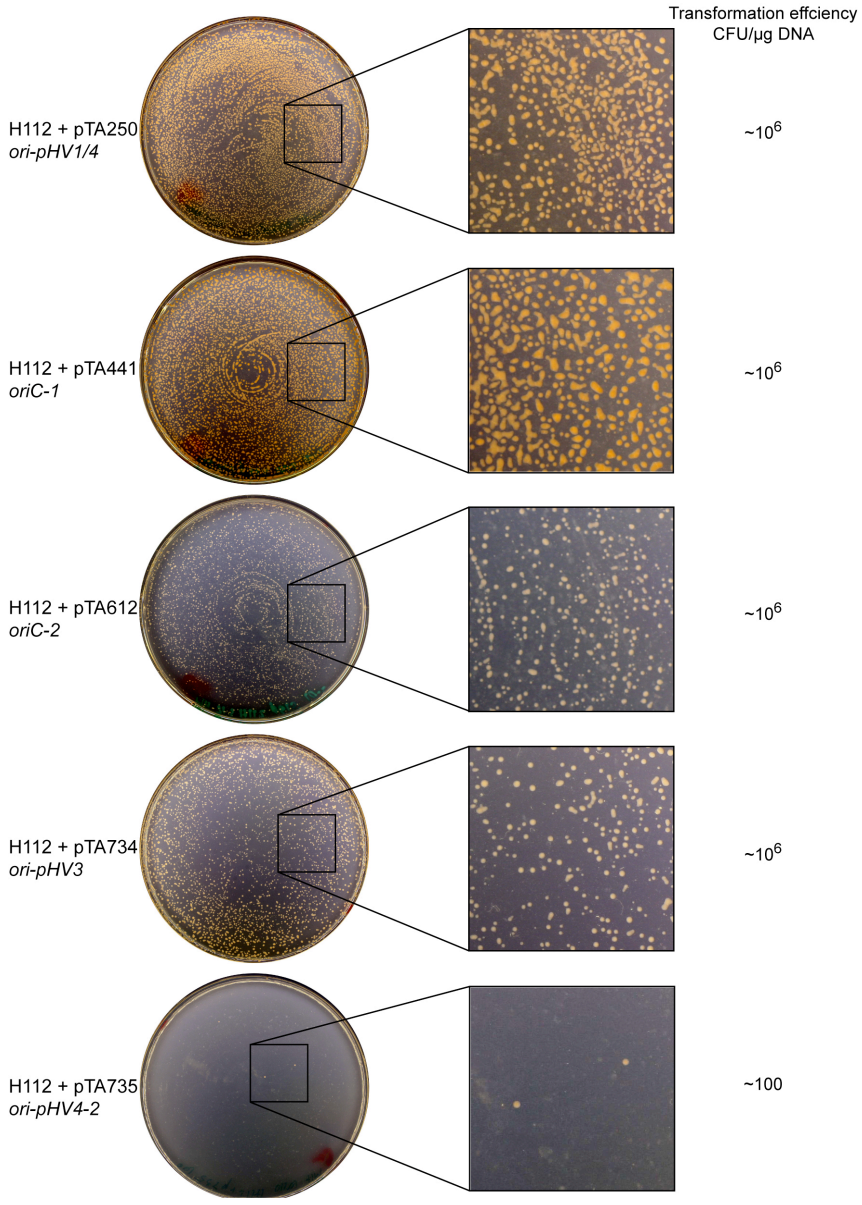


Figure 3.19: Transformation efficiency and colony size of *H. volcanii* origin-based plasmids in the ARS assay

H. volcanii H112 (Δ*pyrE2*, Δ*radA*) transformed with 1 µg of pTA250 (*ori-pHV1/4*), pTA441 (*oriC-1*), pTA612 (*oriC-2*), pCN26/pTA734 (*ori-pHV3*) or pCN27/pTA735 (*ori-pHV4-2)*. Cells were diluted 100-fold and 100 µl was plated on Hv-Ca, except for pCN27/pTA735 where transformants were plated without dilution. Plates were incubated at 45°C for 15 days. CFU = colony forming units

Each *H. volcanii* origin showed a similar transformation efficiency in the ARS assay except for *ori-pHV4-2* which is drastically lower. *oriC-1* and *ori-pHV1/4* transformant colonies are similar in size but *ori-pHV1/4* transformants take ~10 days to produce colonies whereas *oriC-1* transformants take ~7 days to grow to the same size. *oriC-2*, *ori-pHV3* and *ori-pHV4-2* transformants take ~14 days to produce smaller colonies. Colony sizes do not change after propagation on solid media.

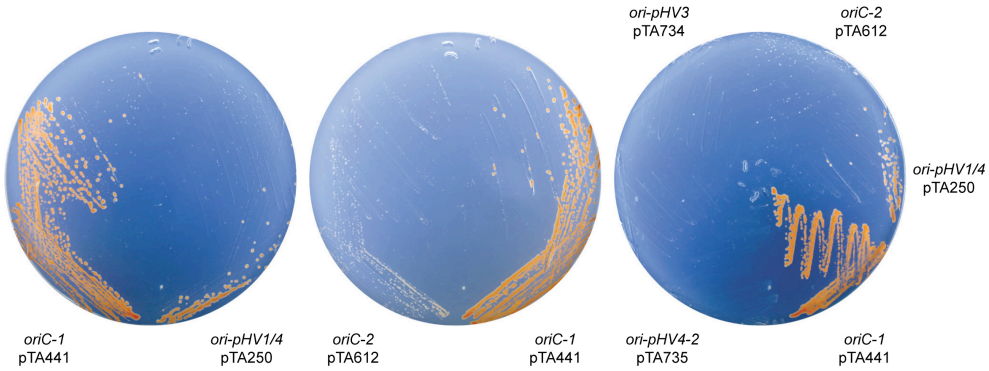


Figure 3.20: Propagation of origin transformants
H112 (Δ*pyrE2*, Δ*radA*) transformed with the origin clones shown and restreaked on selective Hv-Ca plates to show relative colony sizes. Plates incubated at 45°C for 13 days.

3.7.2 Stability of *oriC-2*, *ori-pHV3* and *ori-pHV4-2* ARS plasmids

The stability of *oriC-1* and *ori-pHV1/4*-based plasmids was addressed quantitatively in 3.4.1. Stability of the remaining origins was assessed qualitatively by propagation on non-selective solid media (Hv-YPC). This was followed by restreaking on selective solid media (Hv-Ca). All 6 transformants tested for *oriC-2*, *ori-pHV3* and *ori-pHV4-2* failed to grow on Hv-Ca, indicating complete loss of the origin based plasmid. Therefore in contrast to the *oriC-1* and *ori-pHV1/4* ARS plasmids these plasmids are unstable (compare with Figure 3.11).

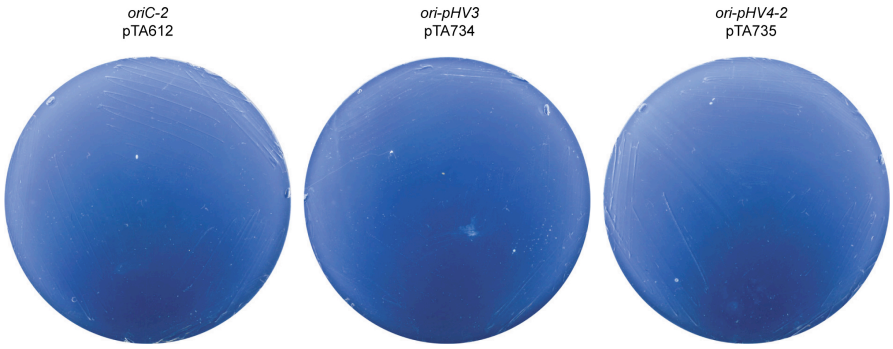


Figure 3.21: Stability of *oriC-2*, *ori-pHV3* and *ori-pHV4-2*-based plasmids
H112 (Δ*pyrE2*, Δ*radA*) transformed with the origin clones shown, streaked on Hv-Ca after propagation without selection on solid Hv-YPC media.

3.7.3 ARS plasmid copy number

To measure the ARS plasmid copy number copy number of origin-based plasmids, total DNA from ARS transformants was displayed on a gel, transferred to a membrane and hybridised with a probe for the origin sequence. The intensity of the episomal plasmid signal was compared to the chromosomal origin locus to get an estimate of relative copy number. Initially a genomic DNA preparation method based on spooling high molecular weight DNA from an ethanol precipitation interface was used. Despite maintaining selection for the ARS plasmid by growing cultures in ura$^-$ media (Hv-Ca broth), the DNA prepared using this method often showed an absence of episomal plasmid. Since the ARS plasmids have a significantly lower molecular weight than genomic DNA it was speculated that they could be preferentially lost during the spooling phase. To optimise the DNA preparation method a Southern blot was carried out to analyse total DNA from H112 (Δ*pyrE2*, Δ*radA*) transformed with pTA441 (*oriC-1*), prepared using either spooling with isopropanol precipitation, centrifugation with isopropanol precipitation, phenol-chloroform purification or a lysis method with no purification step (methods described in 2.2.7). Lysing *H. volcanii* cells in water with no additional purification steps proved to be the DNA preparation method which gave the least variation in chromosomal and plasmid locus band intensities.

Using this method the copy number of pTA441 (*oriC-1*) and pTA612 (*oriC-2*) was assessed.

Table 3.3: *oriC-1* and *oriC-2* ARS plasmid copy number
Each chromosome : plasmid ratio is the mean of six individual ARS transformant ratios.

	pTA441 (*oriC-1*)		**pTA612 (*oriC-2*)**	
Experiment	1	2	1	2
Chromosomal origin locus : episomal origin locus	0.264	1.24	0.551	1.92
Standard error	0.03	0.48	0.07	0.64

This variability could be due to large copy number variations within populations. Since it was so variable and difficult to interpret this method was not pursued further.

One of the few antibiotics that is effective in *H. volcanii* is mevinolin (simvastatin). Mevinolin inhibits HMG-CoA reductase (3-hydroxy-3-methylglutaryl coenzyme A) which results in blocked production of isoprenoid lipids found in the cell membrane (Cabrera *et al.,* 1986). Mevinolin resistance in *H. volcanii* is based on over production of HMG-CoA reductase due to a promoter mutation and has been exploited in genetic systems (Wendoloski *et al.,* 2001). Compared with the amount of *pyrE2* gene product needed to grow on ura$^-$ media, significantly more mutant HMG-CoA reductase is required to confer mevinolin resistance (T. Allers, unpublished observations). Therefore only cells containing plasmids with efficient origin sequences can maintain enough episomal plasmid to produce colonies if mevinolin resistance is the basis for selection. The first shuttle vectors developed for *H. volcanii* were based on the pHV2 plasmid which is present in approximately six copies per chromosome (Charlebois *et al.,* 1987). Shuttle vectors based on the pHV2 origin of replication can be used with the Mev^R marker but plasmids based on *ori-pHV1/4* do not produce transformants (T. Allers, unpublished observations). To get a qualitative measure of copy number, each *H. volcanii* origin and the *oriC-X* control region were subcloned in to a non-replicating vector containing the Mev^R marker (pTA401).

Table 3.4: Mev^R marked origin plasmids

***pyrE2*$^+$ plasmid**	**Insert**	**Restriction digest**	**pTA401 restriction digest**	**Mev^R plasmid**
pTA313	*oriC-1*	*Eag*I	*Not*I	pTA821
pTA250	*ori-pHV1/4*	*Eco*RI/*Xho*I	*Eco*RI/*Xho*I	pTA767
pTA612	*oriC-2*	*Eco*RI/*Hin*dIII	*Eco*RI/*Hin*dIII	pTA765
pTA734	*ori-pHV3*	*Eco*RI/*Not*I	*Eco*RI/*Not*I	pTA787
pTA735	*ori-pHV4-2*	*Eco*RI/*Not*I	*Eco*RI/*Not*I	pTA788
pTA451	*oriC-X*	*Bam*HI/*Xho*I	*Bam*HI/*Xho*I	pTA812

The Mev^R origin plasmids were constructed as detailed in table 3.4 and used to transform H112 (Δ*pyrE2*, Δ*radA*) in an ARS assay. Figure 3.22 shows that only the *oriC-1* mevinolin-based plasmid (pTA821) produced transformants. Therefore only *oriC-1*-based plasmids are maintained at a high enough copy number to produce enough HMG-CoA reductase to allow growth on media containing mevinolin.

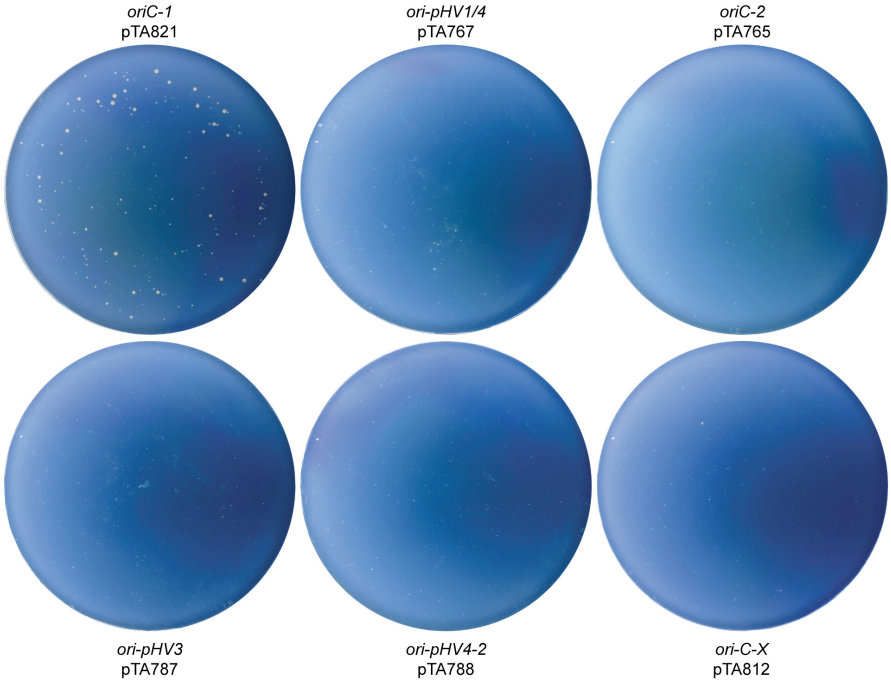


Figure 3.22: ARS plasmid copy number
H112 (Δ*pyrE2*, Δ*radA*) transformed with 1 µg of Mev^R marked origin clones: pTA767 (*ori-pHV1/4*), pTA821 (*oriC-1*), pTA765 (*oriC-2*), pTA787 (*ori-pHV3*), pTA788 (*ori-pHV4-2*) or pTA812 (*oriC-X*). Cells were diluted 100-fold and 100 µl was plated on Hv-YPC supplemented with 4 µg mevinolin. Plates were incubated at 45°C for 15 days.

The characteristics of the *H. volcanii* origins of replication on ARS plasmids are summarised in Table 3.5.

Table 3.5: Characteristics of replication origins

Origin	Observed	Transformant colony size[a]	Transformation efficiency[b]	Plasmid stability	Copy number
oriC-1	Chromosome	Big	~10^6	Maintained	High
oriC-2	Chromosome	Small	~10^6	Lost	Low
ori-pHV1/4	pHV1/pHV4	Big	~10^6	Partial	Low
ori-pHV3	pHV3	Small	~10^6	Lost	Low
ori-pHV4-2	Chromosome	Small	~100	Lost	Low

[a] H112 transformed with ARS plasmid,

[b] Colony-forming units/µg DNA

3.8 Origin deletion

The ARS assays using the intergenic origin regions show that the adjacent *orc* gene is not required *in cis* for origin activity. This is in contrast to previous work in *Halobacterium sp. NRC-1* where the identified origin sequence only functioned effectively in the ARS assay when the region cloned included the adjacent *orc7* gene (Berquist and DasSarma, 2003). In addition to being non-essential *in cis* for origin activity, the *orc* genes of *H. volcanii* associated with *oriC-1*, *oriC-2* and *ori-pHV1/4* (*orc1*, *orc5* and *orc10* respectively) can be deleted individually (Norais, 2007). The Δ*orc1*, Δ*orc5* and Δ*orc1*, Δ*orc10* double deletion strains are also viable so these *orc* genes are also non-essential in their chromosomal context (Norais, 2007).

As described earlier, *ori-pHV1/4* can be deleted with no gross phenotypic effects despite the strong bias for maintenance of the replication origin. The selection pressure for retaining *ori-pHV1/4* was reflected by the 1000-fold outnumbering of wild-type restoration to *ori-pHV1/4* deletion events. Thorsten Allers tested chromosome origin deletion viability by constructing *oriC-1* and *oriC-2* deletion strains. The *Eco*RI-*Bsp*EI fragment of pTA441 containing *oriC-1* was replaced with the *Bam*HI *trpA*$^{+}$ fragment of pTA298 to generate the *oriC-1* deletion construct pTA532. The *Rsr*II fragment of pTA419 containing *oriC-2* was replaced with the *Bam*HI *trpA*$^{+}$ fragment of pTA298 to generate the *oriC-2* deletion construct pTA531.

To confirm that no origin activity was retained after deletion of the DUE and the surrounding ORB elements, the deletion construct clones were tested in the ARS assay. No transformants were recovered when pTA531 and pTA532 were used to transform H112 (Δ*pyrE2*, Δ*radA*) (see Figure 3.23).

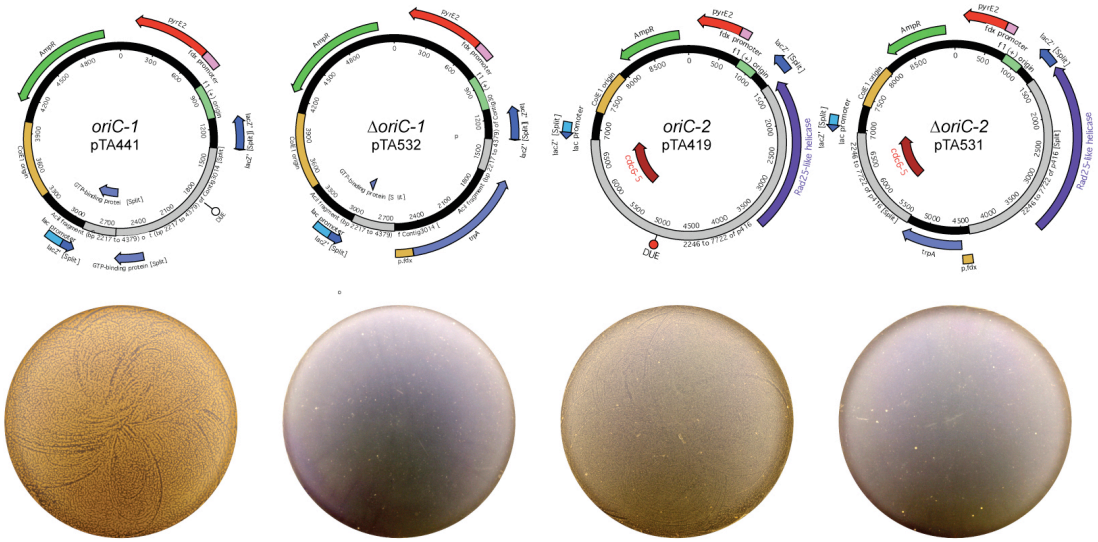


Figure 3.23: *oriC-1* and *oriC-2* deletion construct ARS assay
H112 (Δ*pyrE2*, Δ*radA*) transformed with 1 µg of the *oriC-1* or *oriC-2* intergenic clones and deletion constructs. Transformants were plated undiluted on Hv-Ca and plates were incubated at 45°C for 15 days.

pTA531 and pTA532 were used to transform H53 (Δ*pyrE2*, Δ*trpA*) to prototrophy for uracil and integration at the correct locus was verified using Southern blot analysis (Figure 3.24, panel A, strain H447 *oriC-1*+::[Δ*oriC-1*::*trpA*$^+$ *pyrE2*$^+$] and panel D, strain H446 *oriC-2*+::[Δ*oriC-2*::*trpA*$^+$ *pyrE2*$^+$]). Selection was relieved for uracil to encourage intramolecular recombination and pop-out. Counter-selection with 5-FOA was then used to ensure loss of integrated pTA531 or pTA532. Selection for tryptophan was maintained to directly select for recombination events resulting in the deletion mutant.

It was possible to delete *oriC-1* (H473) and *oriC-2* (H471) individually. This was shown by diagnostic digest and Southern blot analysis (Figure 3.24, panels A and D) and confirmed by using the deleted origin sequence to probe intact DNA displayed on a PFG (Figure 3.24, panels B and C). Putative deletion mutants were plated on Hv-Ca+5-FOA +/- trp and recovery of *oriC-1* and *oriC-2* deletions (*trpA*$^+$, 5-FOA-resistant cells) occurred at a frequency of ~10%. This indicated that there is a much weaker bias for maintenance of *oriC-1* and *oriC-2* than *ori-pHV1/4*.

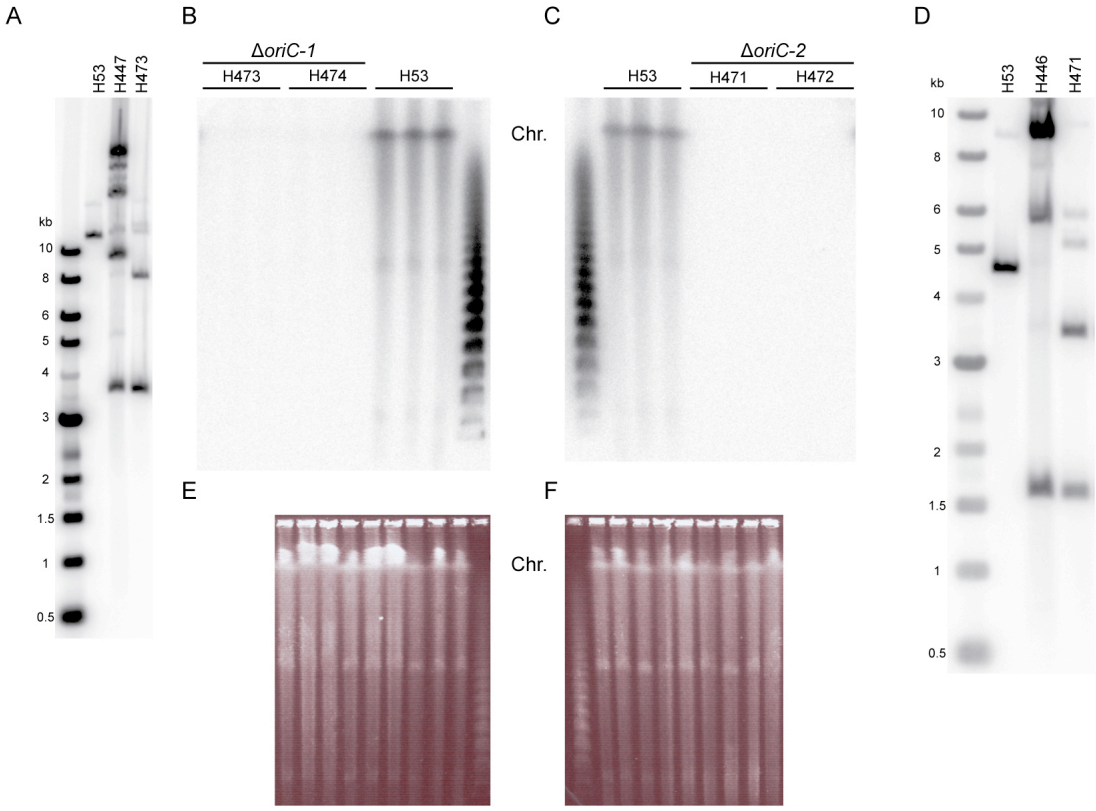


Figure 3.24: *oriC-1* and *oriC-2* deletion

(A) Southern blot showing the wild-type H53 (Δ*pyrE2*, Δ*trpA*), downstream pop-in H447 (Δ*pyrE2*, Δ*trpA*, *oriC-1*+::[Δ*oriC-1*::*trpA*$^+$ *pyrE2*$^+$]) and *oriC-1* deletion strain H473 (Δ*pyrE2*, Δ*trpA*, Δ*oriC-1*::*trpA*$^+$). Genomic DNA was digested with *Stu*I/*Blp*I and probed with pTA313 (*Xho*I/*Stu*I). H473 bands are indicative of Δ*oriC-1*::*trpA*$^+$. Performed by Thorsten Allers.

(B) Southern blot of intact wild-type H53 (Δ*pyrE2*, Δ*trpA*) and Δ*oriC-1* (H473, H474) genomic DNA displayed on a PFG and probed with the deleted *oriC-1* region (pTA441 *Bsp*EI/*Eco*RI).

(C) Southern blot of intact wild-type H53 (Δ*pyrE2*, Δ*trpA*) and Δ*oriC-2* (H471, H472) genomic DNA displayed on a PFG and probed with the deleted *oriC-2* region (pTA419 *Rsr*II).

(D) Southern blot showing the wild-type H53 (Δ*pyrE2*, Δ*trpA*), downstream pop-in H446 (Δ*pyrE2*, Δ*trpA*, *oriC-2*+::[Δ*oriC-2*::*trpA*$^+$ *pyrE2*$^+$]) and *oriC-2* deletion strain H471 (Δ*oriC-2*::*trpA*$^+$). Genomic DNA was digested with *Stu*I/*Blp*I and probed with pTA419 (*Bam*HI/*Not*I). H471 bands are indicative of Δ*oriC-2*::*trpA*$^+$. Performed by Thorsten Allers.

(E) Ethidium bromide staining of intact *H. volcanii* genomic DNA showing presence of the main chromosome in Δ*oriC-1* strains, lanes as in B.

(F) Ethidium bromide staining of intact *H. volcanii* genomic DNA showing presence of the main chromosome in Δ*oriC-2* strains, lanes as in C.

3.9 Genetic screen for origins in a Δ*ori-pHV1/4*, Δ*oriC-1* background

Five origins of replication have been characterised in this work but the techniques used to identify them do not represent an exhaustive origin search. Genetic screens for ARS elements could potentially produce a representative sample of all *H. volcanii* replication origins. However, both genetic screens isolated only a single origin. Predictions based on bioinformatics can guide genetic studies but they exclude many regions by narrowing the focus. Since the genetic screens have not isolated every known replication origin, it is possible that more origins may exist. An example of an inferred but unidentified origin exists on the pHV1 and pHV4 replicons. When *ori-pHV1/4* is deleted the replicons are still present and are therefore using alternative origins. Rather than test each *orc*-associated intergenic region, a third genetic screen for ARS elements was performed. Using the same rationale as in 3.2, Δ*oriC-1* was deleted in a Δ*ori-pHV1/4* background so that an ARS genetic screen could be carried out without domination by these origins. Screening for ARS elements using Δ*oriC-1*, Δ*ori-pHV1/4* genomic DNA could potentially isolate novel origins or the three identified in 3.5.

The *Eco*RI-*Bsp*EI fragment of pTA441 containing *oriC-1* was replaced with the *Bam*HI/*Xba*I *hdrB* fragment from pTA187 to generate the *oriC-1* deletion construct pTA946. Strain H300 (Δ*pyrE2*, Δ*trpA*, Δ*hdrB*, Δ*ori-pHV1/4::trpA*) was transformed with pTA946 (Δ*oriC-1*::*hdrB pyrE2*$^{+}$) and transformants were selected on Hv-Ca. Integration was checked using a *Bsp*EI digest of transformant genomic DNA (Figure 3.25, panel C, H955). Selection was relieved for uracil to encourage intramolecular recombination and pop-out. Counter-selection with 5-FOA was then used to ensure loss of integrated pTA946. Selection for tryptophan was maintained to directly select for recombination events resulting in the deletion mutant. Deletion of *oriC-1* was confirmed using a *Bsp*EI diagnostic digest of genomic DNA (Figure 3.25, panel C, H1023).

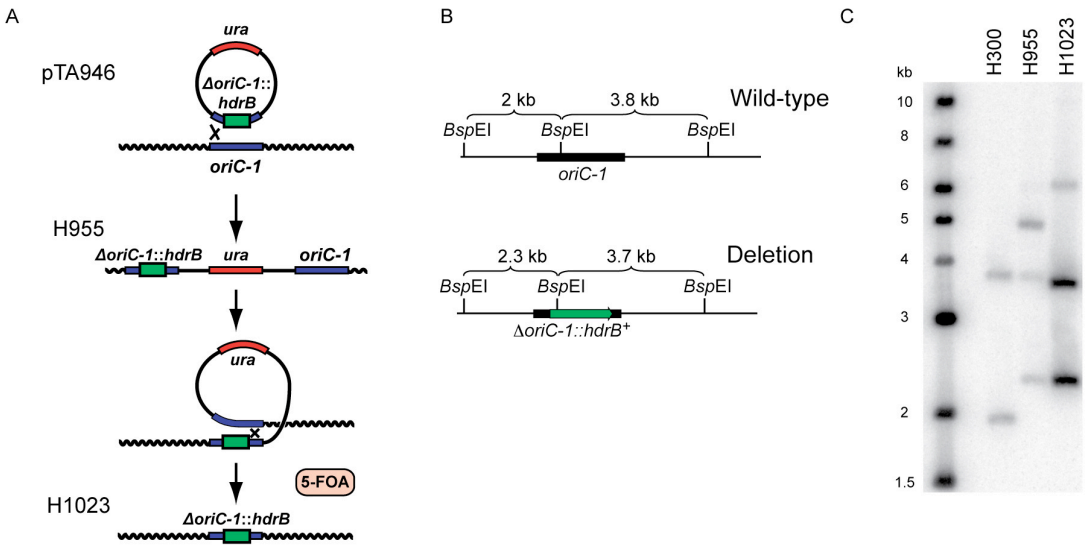


Figure 3.25: *oriC-1* deletion in a Δ*ori-pHV1/4* background
(A) *oriC-1* deletion strategy using an *hdrB* marked deletion construct.
(B) Diagnostic digest and predicted fragments for all stages of strain construction.
(C) Southern blot showing H300 (Δ*pyrE2*, Δ*trpA*, Δ*hdrB*, Δ*ori-pHV1/4::trpA*), downstream pop-in H955 (Δ*pyrE2*, Δ*trpA*, Δ*hdrB*, Δ*ori-pHV1/4::trpA*, Δ*oriC-1*::[Δ*oriC-1*::*hdrB pyrE2*$^{+}$]) and deletion strain H1023 (Δ*pyrE2*, Δ*trpA*, Δ*hdrB*, Δ*ori-pHV1/4::trpA*, Δ*oriC*-1::*hdrB*$^{+}$). Genomic DNA was digested with *Bsp*EI and hybridised with pTA946.

Genomic DNA from strain H1023 (Δ*pyrE2*, Δ*trpA*, Δ*hdrB*, Δ*ori-pHV1/4*::*trpA*, Δ*oriC-1*::*hdrB*) was prepared using the maxiprep method (2.2.8) and 25 µg was partially digested with 0.5 units/µg of *Aci*I for 30 minutes in suboptimal buffer (NEB 1). 4-8 kb fragments were purified and cloned in the compatible *Cla*I site of pTA131 (Δ*pyrE2*). Library coverage was calculated to be 99.9% (~6000 colonies, *H. volcanii* genome ~4 Mb) using the formula:

$$N = \frac{\ln(1-P)}{\ln\left(1-\frac{I}{G}\right)}$$

N = number of colonies
P = probability of sequence being represented in the library
I = insert size
G = genome size (Clarke and Carbon, 1976)

1 µg of the library was used to transform H112 (Δ*pyrE2*, Δ*radA*) to prototrophy for uracil and transformants were selected on Hv-Ca as in previous ARS screens. *H. volcanii* plasmid preparations from six transformants were passaged through *E. coli* and sequenced.

All six transformants contained the same insert, a region on the main chromosome including *orc2*. The intergenic region adjacent to *orc2* contains features characteristic of *H. volcanii* replication origins (DUE, repeats) (Figure 3.26). Two of these plasmids were used to transform H112 (Δ*pyrE2*, Δ*radA*) to confirm that this sequence acted as an ARS element. Transformants were obtained with ~10-fold lower frequency than *ori-pHV1/4*, *oriC-1*, *oriC-2* and *ori-pHV3* clones. Transformant size was large and similar to that seen in *ori-pHV1/4* and *oriC-1* transformants. This origin was named *oriC-3* because the *orc5*-associated origin previously described is established as *oriC-2*.

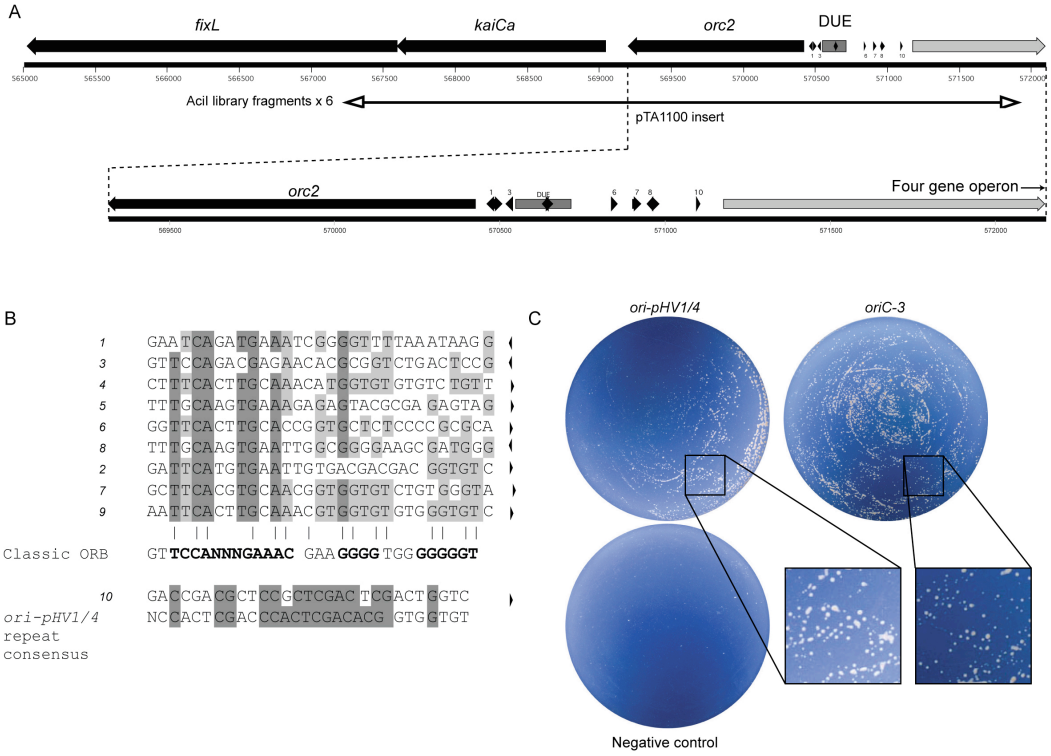


Figure 3.26: Features of *oriC-3*

(A) Sequence features of *oriC-3*. A four gene operon is located downstream of *orc2* (not completely shown). See Figure 3.1 for key.

(B) Sequence of repeats (numbered in A) within the intergenic region adjacent to *orc2*. Repeat orientation is indicated by arrows and conserved positions are shaded.

(C) H112 (Δ*pyrE2*, Δ*radA*) transformed with 1 µg of *oriC-3* (pTA1100), *ori-pHV1/4* (pTA250) and dH_2O. *ori-pHV1/4* transformants were plated with a 100-fold dilution, *oriC-3* transformants with a 10-fold dilution and negative control (dH_2O) transformants were plated undiluted (all on Hv-Ca). Plates were incubated at 45°C for 15 days.

Like other origins identified in this work, *oriC-3* contains several repeats with similarity to the ORB consensus surrounding the DUE. While they all show some similarity to the mini-ORB consensus sequence, the G-string feature of the full length ORB is less well conserved (Figure 3.26, panel B). The repeats are more degenerate than the highly similar repeats at *ori-pHV1/4*, *ori-pHV3* and *ori-pHV4-2*. The DUE is 57% A/T-rich in comparison with 33% for the entire genome. This is similar to the DUE A/T content of *oriC-2* (55%) and *oriC-1* (61% and 56%). Unlike *oriC-1* and *oriC-2*, *oriC-3* is not situated at a nucleotide skew inflection point (Figure 3.27).

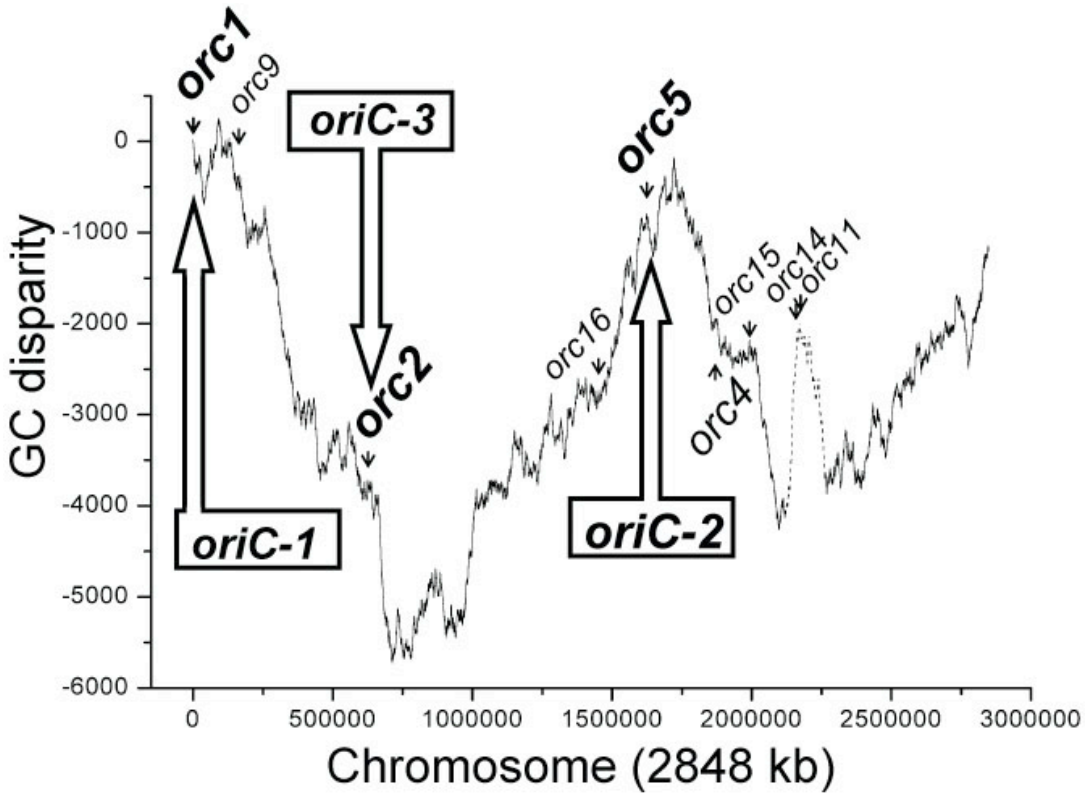


Figure 3.27: GC disparity of main chromosome
Main chromosome G/C skew showing the *orc* gene positions. *orc* genes associated with an origin of replication are bolded and identified origins are marked. *ori-pHV4-2* is also located on the main chromosome in laboratory strains but the location is unknown.

orc2 is adjacent to an operon containing *kaiCa* and *fixL*. In cyanobacteria KaiC is a regulator of circadian rhythms in conjunction with KaiA and KaiB (reviewed in Johnson *et al.,* 2008). KaiC is found in archaea and bacteria and consists of two RecA-like domains joined head to tail (Haldenby *et al.,* 2009). *H. volcanii* contains

four *kaiC* homologues which are named *kaiCa-d*. Like all archaea and most bacteria without circadian rhythms, *H. volcanii* lacks *kaiA* and *kaiB*. FixL is a bacterial histidine kinase that contains a heme-based oxygen-sensing domain. In partnership with FixJ it regulates the transcription of genes involved in nitrogen fixation (Monson *et al.,* 1992). Other heme-based sensors have been linked to circadian rhythms in mice (Gilles-Gonzalez and Gonzalez, 2005). *oriC-3* is also adjacent to an operon of four unidentified ORFs. Gene three (ORF01321) in this operon is a member of a protein family with unknown function and is conserved in archaea and bacteria. The other three genes encode hypothetical proteins. This operon is not conserved across archaea and only exists in *Halorubrum lacusprofundi*, where the first three genes and their order are conserved.

The *oriC-3* intergenic region associated with *orc2* has not yet been tested for origin activity. However, given the sequence features of this region and the similarity to the other five characterised origins, it is likely that the intergenic region can support maintenance of a plasmid in an ARS assay. *oriC-3* is the fourth replication origin to be located on the main chromosome. Since there are discrepancies between the genome sequence and laboratory strains (such as the locations of *ori-pHV4-2* and *ori-pHV1/4*), this needs to be confirmed by PFG analysis.

The locations of *H. volcanii* replication origins identified in this work are summarised in Figure 3.28.

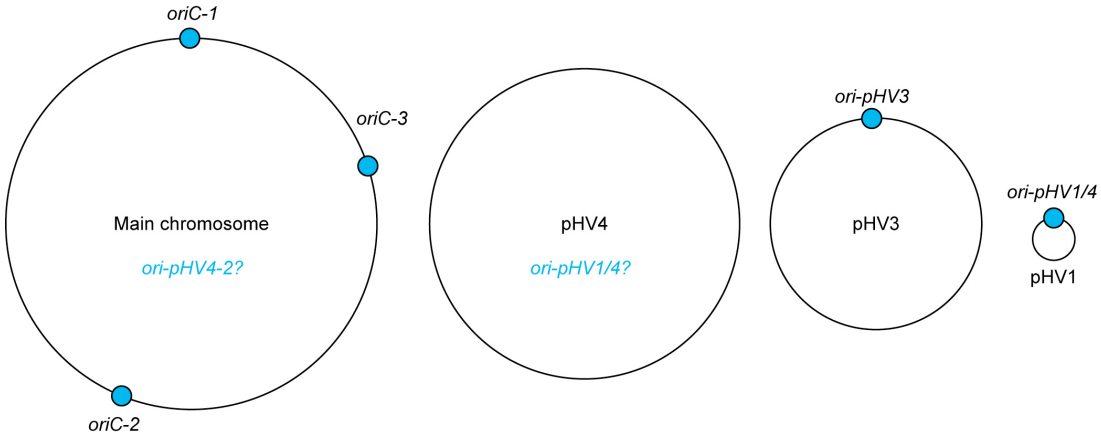


Figure 3.28: Origins of replication in *Haloferax volcanii*
Location of replication origins in *H. volcanii*. Main chromosome is not to scale.

3.10 Discussion

Several origins of replication distributed across the *H. volcanii* genome have been isolated using a combination of genetics, biochemistry and bioinformatics. Multiple origins were located on the main chromosome, providing the first experimentally verified example of multiple origins on a single replicon for a euryarchaeon. The general features of *H. volcanii* replication origins are similar and match characterised replication origins in other archaea. Each origin is adjacent to an *orc1*/*cdc6* gene and consists of an A/T-rich sequence DUE surrounded by repeats. At several origins the repeats show resemblance to the ORB sequence.

The intergenic regions that support ARS activity range in size from 604 bp (*ori-pHV4*-2) to 2.4 kb (*oriC-2*). It is likely that *oriC-2* can be delimited further because the origin features are concentrated within a 1.3 kb region which is similar in size to the *oriC-1* and *ori-pHV1/4* fragments that can support ARS activity. In addition deletion of an 850 bp fragment containing the DUE abrogates ARS activity (see Figure 3.23). The similarities and differences between these origin sequences may have functional relevance *in vivo* for origin-specific binding of replication factors. Multiple conserved repeats at *oriC-1* and *oriC-2* are highly similar to the ORB consensus sequence. In contrast, *ori-pHV1/4*, *ori-pHV3* and *ori-pHV4-2* each contain a different group of repeats which show significant differences from the ORB consensus sequence. The family of repeats at *oriC-3* is less well conserved but also has similarities to the ORB consensus sequence.

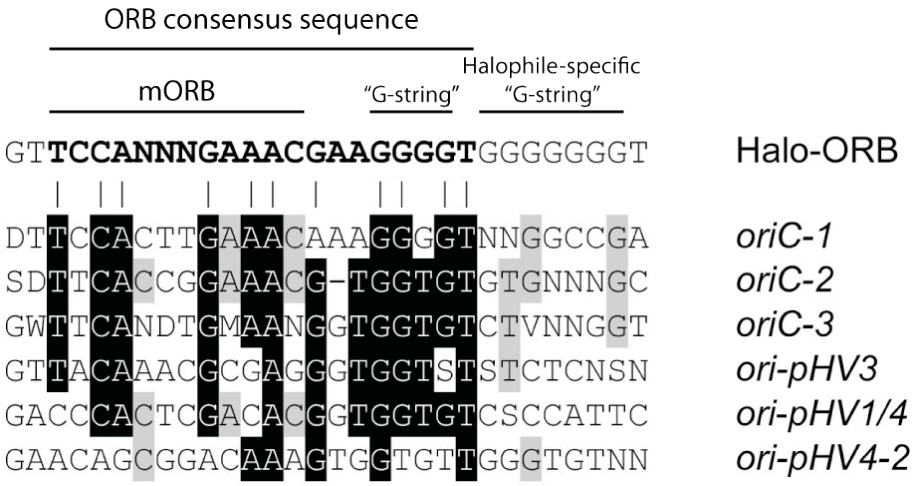


Figure 3.29: *Haloferax volcanii* replication origin repeats
Alignment of the consensus sequences for *H. volcanii* origin repeats with the ORB consensus sequence. For example the *oriC-1* sequence shown is the consensus of the 8 ORB at this locus. The ORB consensus sequence common to all archaea is bolded.

In other archaea the Cdc6/Orc1 initiator protein has been shown to bind directly to ORB elements. Variation of repeat motifs in a species with multiple *cdc6/orc1* homologues may reflect specific binding for different Cdc6/Orc1 proteins. Origin activity does not require the presence of all the ORB sequences at a specific origin because the minimal *oriC-1* (pCN11) sequence excludes the ORB-like repeats numbered 5-8 (Figure 3.4) and still functions in the ARS assay. Each origin except for *oriC-2* has a pair of inverted repeats either side of a DUE. This may be important for directional Cdc6/Orc1 positioning and subsequent MCM recruitment as discussed in 1.5.2. Perhaps the lack of this feature in *oriC-2* is an explanation for the apparent low efficiency of this origin. The DUE motifs contain the initiation point determined by RIP mapping for *oriC-1*, *oriC-2* and *ori-pHV3* whereas at *ori-pHV1/4* initiation is relatively distant from the DUE. The transition point for each strand at *ori-pHV4-2* is slightly offset with one within the DUE and one located outside. Only at *oriC-2* does the initiation point not coincide with a region where the DNA is predicted to be bent. DNA bending and the associated duplex opening may facilitate initiator protein binding (Gimenes *et al.,* 2008) and this could be another reason for the inefficiency of *oriC-2*.

Parallels between initiators and replicators in bacteria and archaea make it tempting to hypothesise that in species with multiple *cdc6/orc1* homologues it is the adjacent *cdc6/orc1* gene product that binds each replication origin and acts as the initiator. The more complex reality is that origin binding is not exclusive to individual Cdc6/Orc1 proteins. This is demonstrated by *S. solfataricus* where origins are recognised by subsets of the three Orc proteins (Robinson *et al.,* 2004). The large number of *cdc6/orc1* homologues in halophiles relative to other archaea is probably related to their dynamic genomes. Horizontal gene transfer, gene duplication and mobile genetic elements have all significantly contributed to halophile genome organisation. This is reflected by the existence of multiple homologues for several genes. In addition to the numerous *cdc6/orc1* homologous, halophiles also contain several copies of genes homologous to the basal transcription factors TBP and TFB (Soppa, 2006). Multiple Orc1/Cdc6 proteins provide scope for origin regulation. For example it has been speculated that multiple Cdc6/Orc1 proteins bind cooperatively to origins and that *cdc6-2* in *S. solfataricus* acts as a negative regulator of replication. In *H. volcanii* there are more *cdc6/orc1* genes than replication origins because the intergenic region adjacent to *orc4* does not function as a replication origin in an ARS

assay. Construction of *cdc6/orc1* deletion strains have shown that not every *cdc6/orc1* homologue is essential (Norais, 2007), therefore their functions may partially overlap. Another potential source of regulation is the interplay between transcription and replication. Five of the six origins identified in this work (*ori-pHV3* is the exception) are situated between divergently transcribed genes. This is also the case for the majority of archaeal replication origins and could indicate that transcriptional interference is relevant to origin firing. Replication factor access to DNA could be enhanced by increasing negative supercoiling due to divergent transcription (Liu and Wang, 1987). This may have been a factor in the evolution of origin placement. Another explanation is that orienting RNA polymerase activity away from replication origin prevents replication and transcription forks colliding. For example the highly transcribed *rrnA* and *rrnB* operons are both oriented away from *oriC-1* and *oriC-2* respectively. Potential for differential regulation of origin activity is particularly important for *H. volcanii* because it needs to coordinate replication of several replicons as well as regulating chromosome copy number.

When *H. volcanii* origins were used to maintain an episomal plasmid in ARS assays they differed in the related properties of transformation efficiency, transformant colony size, plasmid copy number and plasmid stability. However, the relevance of the ARS assay to origin function *in vivo* may be limited. When origins are in their chromosomal context they are subject to different influences than when located on an artificial plasmid. The topology of small artificial plasmids made from mostly foreign DNA will differ drastically from natively packaged chromosomes. Differences in DNA supercoiling due to packaging and transcription have been shown to affect *E. coli oriC* function when it is cloned on a plasmid (Asai *et al.,* 1998). Another consideration is that origin-binding regulatory proteins might require sequences outside the minimal origin region to bind correctly. A more technical caveat is that laboratory culture conditions do not accurately represent the variability of the natural habitat of *H. volcanii*. These limiting factors mean that ARS activity might not directly translate to *in vivo* origin efficiency. The clear nucleotide skew inflection point at *oriC-2* meant that it was expected to function in a similar manner to *oriC-1*. The low efficiency of *oriC-2* ARS plasmids might be demonstrating the limitations of the ARS assay. This highlights the importance of developing a quantitative *in vivo* origin activity assay. RIP mapping can only give a binary indication of origin function.

Based on the original genetic screens and co-transformation results, a hierarchy of origins arising due to competition for replication factors was proposed (Norais *et al.,* 2007). Further co-transformation experiments described here complicate this hypothesis because the "dominant origin" changes depending upon the origin clone used. This would fit the hypothesis if sequences outside the intergenic origin regions were involved in recruiting replication factors. However, the equivalency seen in co-transformations using differentially sized clones of the same origin casts doubt on this. Of the thirteen ARS elements isolated from the initial genetic screens, twelve contained full length *orc10*. This could be a technical artefact due to site preferences of the restriction enzymes used or it is conceivable that *orc10* could be a negative regulator of origin activity for origins other than *ori-pHV1/4*. This would explain all the co-transformation results except for the apparent dominance of intergenic *oriC-1* over the library *ori-pHV1/4* clone. Co-transformation experiments are superficially easy but there are a number of confounding experimental factors that are difficult to control. For example, cells will be transformed by variable numbers and ratios of plasmids. The proportion of each event as well as variables of interest such as plasmid stability and origin efficiency will all contribute to the property of "dominance" seen in the transformant population. Despite the simplicity of the co-transformation concept, the difficulty in controlling all the variables and its ARS assay foundation (limitations discussed above), mean it can be taken no further. However, the co-transformation results have led to the hypothesis that *orc10* is a negative regulator of replication and this can be tested by other means.

Explanations for the isolation of a single origin of replication with each genetic screen are an interesting matter for speculation. The presence of *ori-pHV1/4* on two replicons means that, depending upon the relative copy number of the different *H. volcanii* chromosomes, *ori-pHV1/4* could be over-represented. This could explain its apparent dominance in genetic screens for ARS elements, as could the hypothesis that *orc10* is a negative regulator of replication at other origins. The proliferation of *cdc6/orc1* genes in the halophilic lineage was an opportunity for divergence. Therefore it is possible that the multiple origins of replication on the main chromosome of *H. volcanii* are conditional. The wide growth condition tolerances of *H. volcanii* means there could be an advantage to specialised Cdc6/Orc1 proteins, for example function could vary depending upon salt concentration or temperature. In

the relatively static laboratory culture conditions perhaps this specialisation leads to the isolation of single replication origins using genetic screens.

The most interesting aspect of *H. volcanii* origin distribution is that multiple origins have been found on the main chromosome while only a single origin has been identified for each of the other replicons. It is not known whether each replication origin on the main chromosome fires in every cell cycle. The similar nucleotide disparity inflection points for *oriC-1* and *oriC-2* mean that these origins both likely to be used. Despite the similar bioinformatic characteristics of *oriC-1* and *oriC-2*, these origins differed in their ARS activity (see Table 3.4). This could be due to the aforementioned limitations of the ARS assay or it might suggest that these origins are conditional and that *oriC-2* does not fire efficiently in laboratory conditions. *oriC-3* is not situated at a nucleotide disparity inflection point but functions in an ARS assay with an intermediate transformation efficiency. It should be noted that *oriC-3 in vivo* function and location has not yet been confirmed. The lack of inflection points at the *oriC-3* locus could mean that *oriC-3* has recently translocated to this location. A non-mutually exclusive possibility is that this origin is conditional. The pHV4 skew inflection point for *ori-pHV4-2* origin is clear but pulsed-field gel analysis of laboratory strains locates *ori-pHV4-2* on the main chromosome rather than pHV4. Since the chromosomal coordinates of *ori-pHV4-2* are unknown it is difficult to make a definite statement about nucleotide bias at this origin. Excluding *oriC-3*, each of the *H. volcanii* replication origins functions in its chromosomal context and is situated at a nucleotide skew disparity inflection point. However, their functionality in the ARS assay differs. This was shown by measurements of transformation efficiency and plasmid stability (see Table 3.4). The most severe discrepancy between the genetic assays and other experimental approaches is for *ori-pHV4-2*. The drastically lower transformation efficiency of *ori-pHV4-2* in the ARS assay compared to other origins of replication identified in this work could indicate that *ori-pHV4-2* is an inefficient origin or that it is limited by the ARS assay and/or laboratory conditions. The transformation frequency of the *ori-pHV4-2* ARS plasmid is so low that there are grounds for it being classed as a different type of replication origin. This could be related to the translocation of *ori-pHV4-2* to the main chromosome from pHV4. Given the disagreement between RIP mapping and ARS assay tests of *ori-pHV4-2* origin activity it would be valuable to measure origin function by a third method such as marker frequency analysis. Further development

of experimental methods for *H. volcanii* would be required to accomplish this. In summary the evidence for *oriC-3* and *ori-pHV4-2* presented here is still inconclusive.

The Z-curve analysis of *H. volcanii* replicons is not straightforward compared to other archaeal species. Halophiles present inverted skew polarities when compared to bacteria or thermophilic archaea, which means they contain an excess of Cs on the leading strand of replication. The *H. volcanii* skews demonstrate more fine-scale variation than nucleotide skews for other haloarchaea, therefore the inflection points are not always clearly delineated. Sometimes peaks can be explained by non-origin means, a good example is the peak associated with *orc11*/*14*. This is probably the result of a prophage integration because this 50 kb region has a lower G/C content than the genome average and contains putative viral genes. Skew interpretation can be complicated by multiple origins of replication and dynamic genomes. All skews are a record of the evolutionary pressures on genome regions. Skews measure the long-term accumulation of strand bias so a recently translocated origin will leave an origin signal at its original location. The new origin site however, will not show any nucleotide bias until enough time has elapsed for drift to take effect. The *H. volcanii* genome is relatively dynamic because *ori-pHV1/4* and *ori-pHV4-2* are located on different chromosomes in laboratory strains compared to the DS2 genome sequence. pHV1 is predicted to contain *ori-pHV1/4* but PFG analysis locates it on both the pHV1 and pHV4 replicons. *ori-pHV4-2* is predicted to be situated on pHV4 but in laboratory strains it is present on the main chromosome. These genome rearrangements are not exclusive to origins because the *bgaHv* gene encoding the enzyme *β*-galactosidase is found on the chromosome in laboratory strains but is located on pHV4 in DS2 strains and the genome sequence (see Appendix 7.2). All these noted rearrangements involve pHV4. pHV4 contains many sequences related to mobile genetic elements such as the ISH51 and IS4 insertion sequences. Insertion sequences can trigger genome rearrangements and are common in archaeal genomes (De Palmenaer *et al.,* 2008). Since pHV4 has a drastically different codon bias to the main chromosome (S. Haldenby and S. Delmas, unpublished observations) it is reasonable to suggest that mobile elements significantly contributed to its organisation. Active mobile elements could explain why strain discrepancies all involve pHV4.

It is difficult to speculate about the consequences of multiple chromosomal origins for replication termination because the origin activity profiles are unknown. Termination could rely on fork-collision and resolution, active termination sites or a combination of these strategies. No homologues of the *E. coli* termination region have been found in archaea but this is unsurprising given how species-specific these sequences are within the bacterial domain. If termination always occurs in the same region then this might be evident from nucleotide skew analysis, but as already discussed, the bioinformatic analysis of the *H. volcanii* genome is difficult to interpret. Due to the experimental systems currently established for *H. volcanii*, termination is more difficult to study than replication origins. It is not known if all chromosomal origins fire in every cell cycle or if there is differential regulation. If every origin fires then multiple sets of colliding forks would need to be resolved for each round of replication. If the number of origins that fire changes *in vivo* between cell cycles the termination strategy used must be flexible to cope with the different possibilities and ensure that all regions are duplicated in an appropriate time frame.

The smaller replicons of *H. volcanii* have historically been called plasmids. Based on their replication by Orc-associated origins it was proposed that pHV4, pHV3 and pHV1 should be considered secondary chromosomes (Norais *et al.,* 2007). In addition pHV4 contains predicted essential genes and when pHV3 is lost *H. volcanii* exhibits a significant growth defect and filaments (Wendoloski *et al.,* 2001). It is not known if pHV1 is essential. Given that the major chromosome contains multiple origins of replication the possibility of multiple origins on these secondary chromosomes should also be examined. There are no further obvious skew inflection points associated with pHV4, pHV3 and pHV1 *cdc6/orc1* genes. However, *oriC-3* demonstrates that this is not a requirement. Since origins have been found in other species without *cdc6/orc1* homologues in the vicinity this does not preclude the possibility that there are more origins of replication to be discovered. At the very least there are alternative origins on pHV4 and pHV1 because the deletion of *ori-pHV1/4* does not result in the loss of these replicons. Without candidate origins from bioinformatic analysis the best strategy to find additional origins of replication is to proceed with genetic screens with strains deleted for the "dominant" origins.

Chapter 4: *radA*-independent recombination

4.1 Conditional origins of replication

H. volcanii replication origins could be differentially regulated. *H. volcanii* has a wide tolerance for environmental salt concentration (1-5 M NaCl) so the hypothesis that origin activity changed according to salt conditions was investigated.

A conceivable salt-dependent method of regulating origin activity is that origins of replication are bound by different Cdc6/Orc1 proteins which function optimally at different salt concentrations. High salt concentrations increase the surface tension of water which leads to protein aggregation and precipitation (Danson and Hough, 1997). Highly acidic proteins function better in high salt conditions because acidic residues help prevent protein aggregation by increasing hydration levels (Frolow *et al.,* 1996; Elcock and McCammon, 1998). Halophiles have high intracellular salt concentrations and correspondingly show a bias for acidic amino acids (Paul *et al.,* 2008). pI describes the pH where the estimated charge of a protein is zero and can be estimated from the amino acid sequence, the lower the pI the more acidic the protein. The theoretical pI values of the *H. volcanii* Orc proteins were calculated and are shown in Table 4.1. The predicted pI values of the 16 Orc proteins vary from 4.02 to 5.64 and the origin associated Orc proteins are distributed evenly across this range. The origins recovered from genetic screens (*oriC-1*, *oriC*-3 and *ori-pHV1/4*) are associated with the Orc proteins predicted to be the most acidic which suggested that the *H. volcanii* Orc proteins might function optimally under different salt conditions. In the absence of a biochemical system for measuring Orc-origin binding, the effect of salt concentration on Orc activity was measured indirectly using the ARS assay.

The salt tolerance thresholds for *H. volcanii* are 1 M NaCl and 5 M NaCl but optimal growth occurs in 2.5 M NaCl media (Figure 4.1). Solid and liquid media with a NaCl concentration of 2.05 M (16.3% total salt), 2.5 M (18% total salt) or 2.87 M (21.1% total salt) was prepared. Throughout this chapter these concentrations are referred to as low, normal and high-salt respectively. H112 (Δ*radA*, Δ*pyrE2*) was streaked from frozen stocks on to low, normal and high-salt Hv-YPC media and propagated for a month on Hv-YPC media of the appropriate salt concentration. This period of salt acclimatisation was carried out to prevent salt shock from affecting the ARS assay.

Table 4.1: *Haloferax volcanii* Orc pI values

H. volcanii orc genes ranked by predicted protein product pI.

Gene	Replicon	Associated origin	Experimentally determined origin location	pI
orc1	Chromosome	*oriC-1*	Chromosome	4.02
orc2	Chromosome	*oriC-3*		4.34
orc13	pHV4			4.56
orc10	pHV1	*ori-pHV1/4*	pHV1 and pHV4	4.59
orc7	pHV4			4.63
orc3	pHV4	*ori-pHV4-2*	Chromosome	4.64
orc8	pHV1			4.66
orc9	Chromosome			4.71
orc12	pHV4			4.73
orc5	Chromosome	*oriC-2*	Chromosome	4.83
orc11	Chromosome			4.91
orc4	Chromosome			4.97
orc6	pHV3	*ori-pHV3*	pHV3	4.98
orc16	Chromosome			5.14
orc14	Chromosome			5.42
orc15	Chromosome			5.64

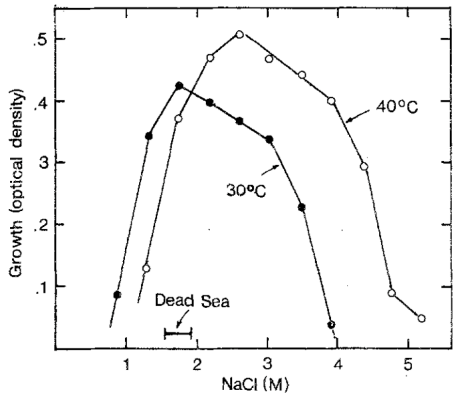


Figure 4.1: Effect of NaCl concentration on *Haloferax volcanii* growth

NaCl concentration varied in tryptone yeast autolysate salt media. Final growth yield was measured when the fastest growing culture approached the stationary phase. From Mullakhanbhai and Larsen, 1975.

Salt-acclimatised H112 (Δ*radA*, Δ*pyrE2*) cells were visualised by microscopy to check for any gross physiological changes resulting from growth in sub-optimal media (Figure 4.2).

Overall cell morphology did not change drastically when the salt concentration of the growth media was altered. Δ*radA* strains of *H. volcanii* are deficient in homologous recombination so DNA repair is impaired (Woods and Dyall-Smith, 1997), in H112 (Δ*radA*, Δ*pyrE2*) populations ~50% of cells are dead and compaction of DNA is common (S. Delmas, unpublished observations). Figure 4.2 shows that DNA compaction and dead cells are seen across the range of salt concentrations. This was observed at the same approximate frequency for each salt concentration. This suggests that there were no gross differences due to growth in sub-optimal media that might affect an ARS assay.

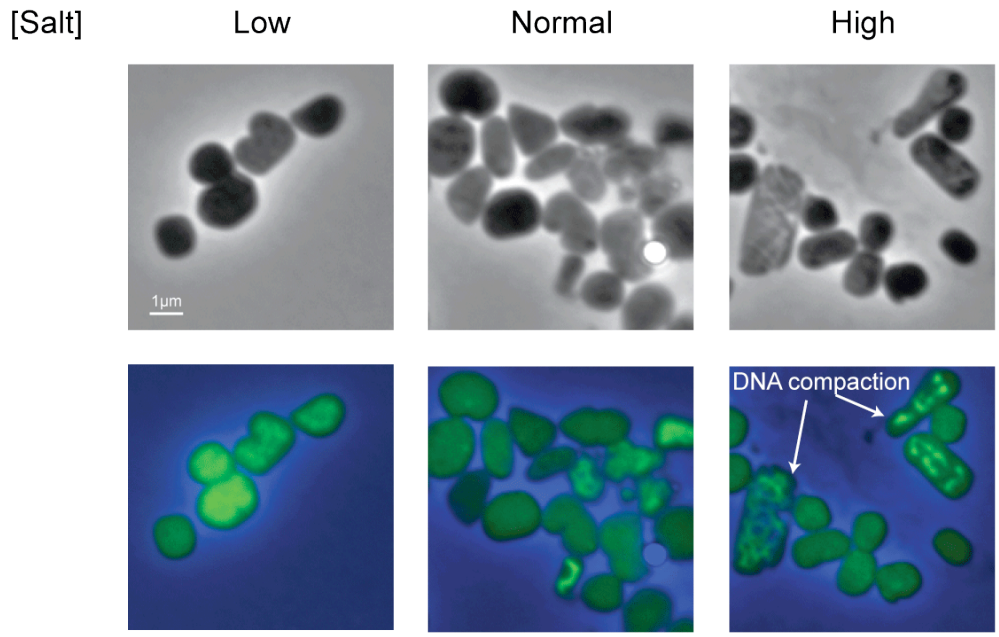


Figure 4.2: *Haloferax volcanii* grown in a range of salt conditions
H112 (Δ*radA*, Δ*pyrE2*) grown for 16 hours in Hv-YPC broth containing either 2.05 M (low-salt), 2.5 M (normal-salt) or 2.87 M (high-salt) NaCl. Cultures were grown to exponential phase (OD_{650} = 0.5) and stained with acridine orange before visualisation. Top row shows bright phase images and the bottom row shows merged yellow fluorescent protein and cyan fluorescent protein filters.

Salt-acclimatised H112 (Δ*radA*, Δ*pyrE2*) was transformed with the intergenic *H. volcanii* origin clones detailed in the Figure legend for Figure 4.3 and 4.4. *oriC-3* was not tested because it had not been identified at this time. The transformations

were carried out three times using transformation solutions and selective Hv-Ca plates with the appropriate salt concentration. The results were more variable than transformations on normal media (compare Figure 4.3 and 4.4), but some generalisations could be made.

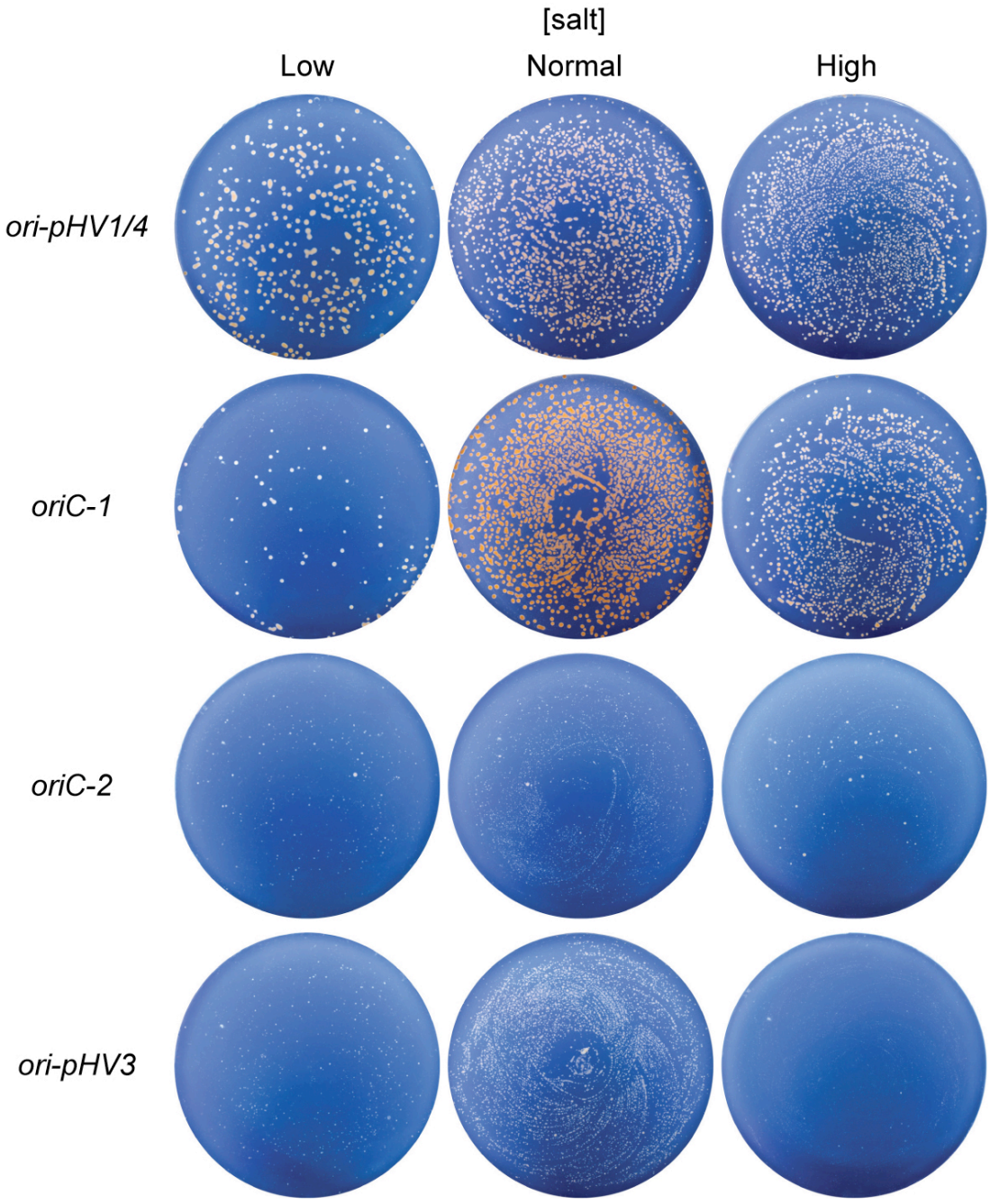


Figure 4.3: ARS assay performed in different salt conditions-1
H. volcanii H112 (Δ*pyrE2*, Δ*radA*) transformed with 1 µg of pTA250 (*ori-pHV1/4*), pTA441 (*oriC1*), pTA612 (*oriC2*) or pTA734 (*ori-pHV3*). Cells were diluted 100-fold and 100 µl of transformants were plated on Hv-Ca plates containing 2.05 M (low-salt), 2.5 M (normal-salt) or 2.87 M (high-salt) NaCl. Plates were incubated at 45°C for 12 days.

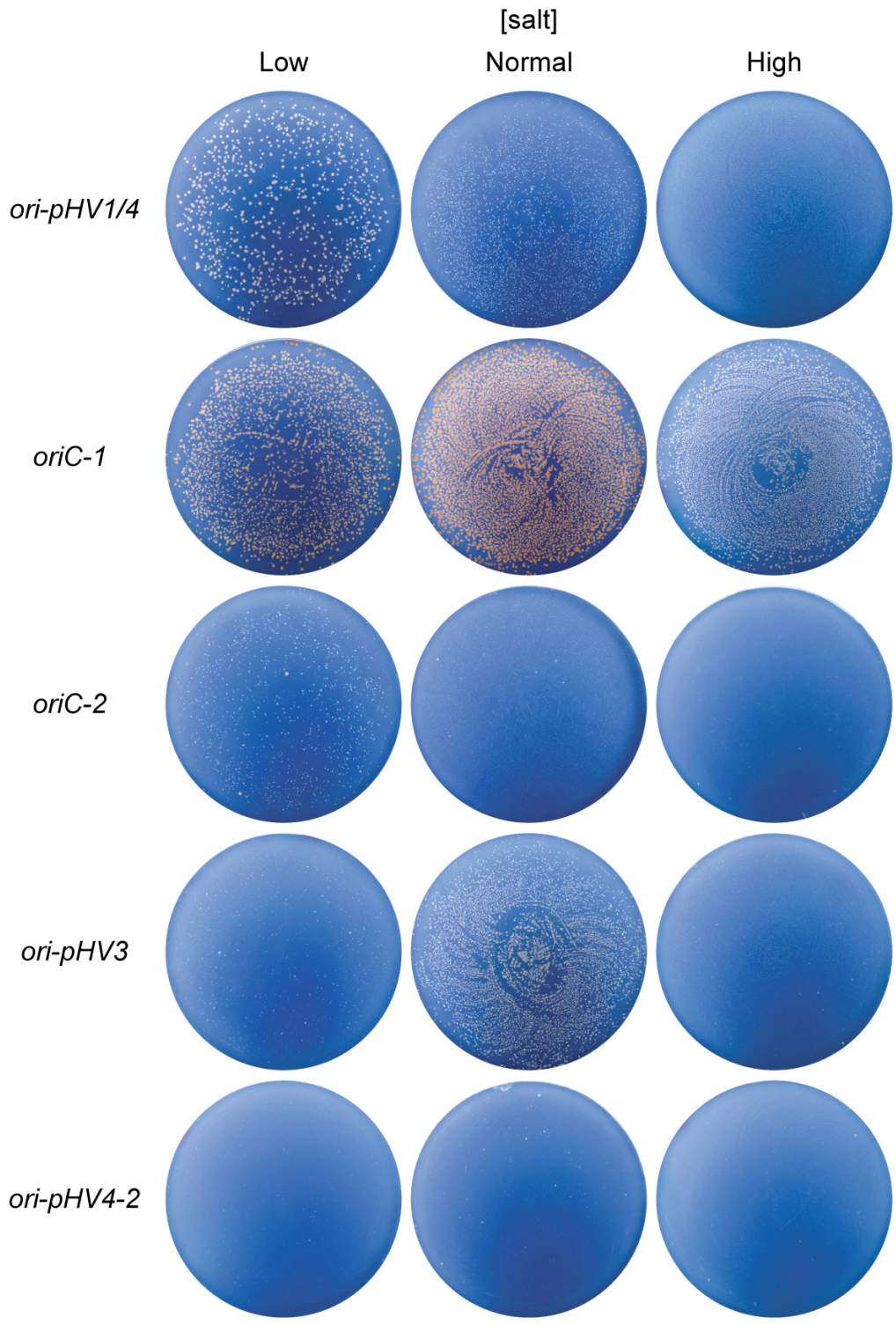


Figure 4.4: ARS assay performed in different salt conditions-2
H. volcanii H112 (Δ*pyrE2*, Δ*radA*) transformed with 1 µg of pTA250 (*ori-pHV1/4*), pTA441 (*oriC1*), pTA612 (*oriC2*), pTA734 (*ori-pHV3*) or pTA735 (*ori-pHV4-2)*. Cells were diluted 100-fold except for pTA735 transformants which were plated without dilution. 100 µl of transformants were plated on Hv-Ca plates containing 2.05 M (low-salt), 2.5 M (normal-salt) or 2.87 M (high-salt) NaCl. Plates were incubated at 45°C for 12 days.

ori-pHV1/4 transformants colonies appeared faster on low salt media but the transformation efficiency was reduced. Colony formation speed varied on high-salt media but colony size was equivalent at all three salt concentrations after 15-17 days of growth. *oriC-1* transformation efficiency was also reduced on low-salt media while in high-salt conditions colonies were smaller and the transformation efficiency varied. *oriC-2* transformants appeared faster on low and high-salt media and produced larger colonies compared with normal-salt transformants. *ori-pHV3* transformants only grew well in optimal salt conditions and no *ori-pHV4-2* transformants were observed at the altered salt concentrations (data from one experiment only). All the colony size differences remained after restreaking transformants on media containing the appropriate salt concentration.

ori-pHV1/4 and *oriC-1* transformants grown in different salt conditions were examined for episomal plasmid maintenance. Episomal maintenance of the other origin-based plasmids was not examined due to time constraints and the length of time required to grow *oriC-2* and *ori-pHV3* transformants in selective broth. Crude DNA was prepared from transformants grown in selective broth (Hv-Ca at the appropriate salt concentration) and digested to linearise the transforming plasmid using *Eco*RI for *ori-pHV1/4* transformants and *Bsp*EI for *oriC-1* transformants

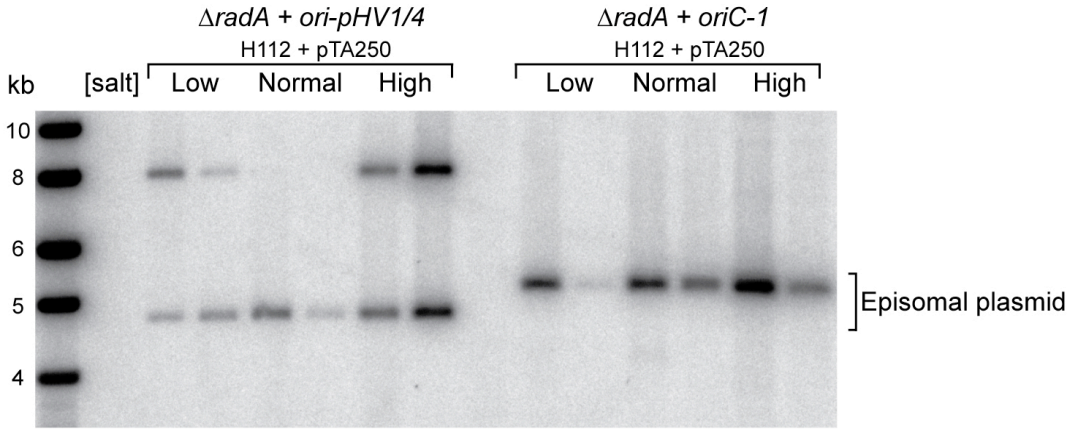


Figure 4.5: Episomal plasmid maintenance in different salt conditions
Southern blot of crude DNA preparations from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*) or pTA441 (*oriC-1*). DNA was digested with *Eco*RI for pTA250 (4.7 kb) and *Bsp*EI for pTA441 (5.1 kb) transformants. The blot was probed with *pyrE2* (pTA52 *Nco*I/*Xba*I).

The *ori-pHV1/4* and *oriC-1* plasmids linearised to the correct size and were therefore replicating episomally. An unidentified band of 8 kb was present in the low and high-salt *ori-pHV1/4* transformants. The *pyrE2* probe used should only hybridise with the *pyrE2* gene on the plasmid because *pyrE2* is deleted in H112. No other *H. volcanii* sequences have significant homology to the probe.

4.2 Identification of the 8 kb band

For the rest of this chapter investigation of the unexpected 8 kb band will be described. The episomal plasmid analysis was broadened to include more transformants in order to determine how common this feature was (Figure 4.6). Twelve *ori-pHV1/4* and *oriC-1* transformants were tested at each salt concentration and the unidentified band was only present in *ori-pHV1/4* transformations performed in high-salt conditions. This event occurred at a frequency of 17% (2/12). The recombination deficiency of H112 meant that the 8 kb band was not the product of an integration event, therefore a variety of strategies were employed to identify it.

To identify the 8 kb band, an *Eco*RI DNA library was made from a H112 pTA250 (*ori-pHV1/4*) high salt transformant that presented the unexpected band (Figure 4.6, transformant 9). Fragments of ~8 kb were purified after agarose gel electrophoresis separation and cloned in *Eco*RI digested pBluescript II SK+. Clones were patched out and colony lifts from the patched plates were hybridised with the same probe that exposed the 8kb band (*pyrE2*, pTA52 *Nco*I/*Xba*I). Plasmid from a clone that hybridised with *pyrE2* was isolated (pTA894) and the sequenced insert was found to correspond to *ori-pHV1/4*.

ori-pHV1/4 is located on a 7.8 kb *Eco*RI fragment of pHV1. However, this fragment shares no homology with *pyrE2* and therefore should not be hybridising with this probe. A second *EcoRI* DNA library was made from a H112 pTA250 (*ori-pHV1/4*) high-salt transformant that presented the unexpected band. Fragments of ~8 kb were purified from an agarose gel, self-ligated and used to transform *E.coli*. Transformants were recovered, which meant that at least the ColE1 replication origin and ampicillin resistance marker were present in the 8 kb fragment library. Since an *ori-pHV1/4*-based plasmid (pTA250) was used to transform strains presenting the unidentified 8 kb band, a variety of pTA250 fragments were used as Southern blot probes to locate all the plasmid sequences.

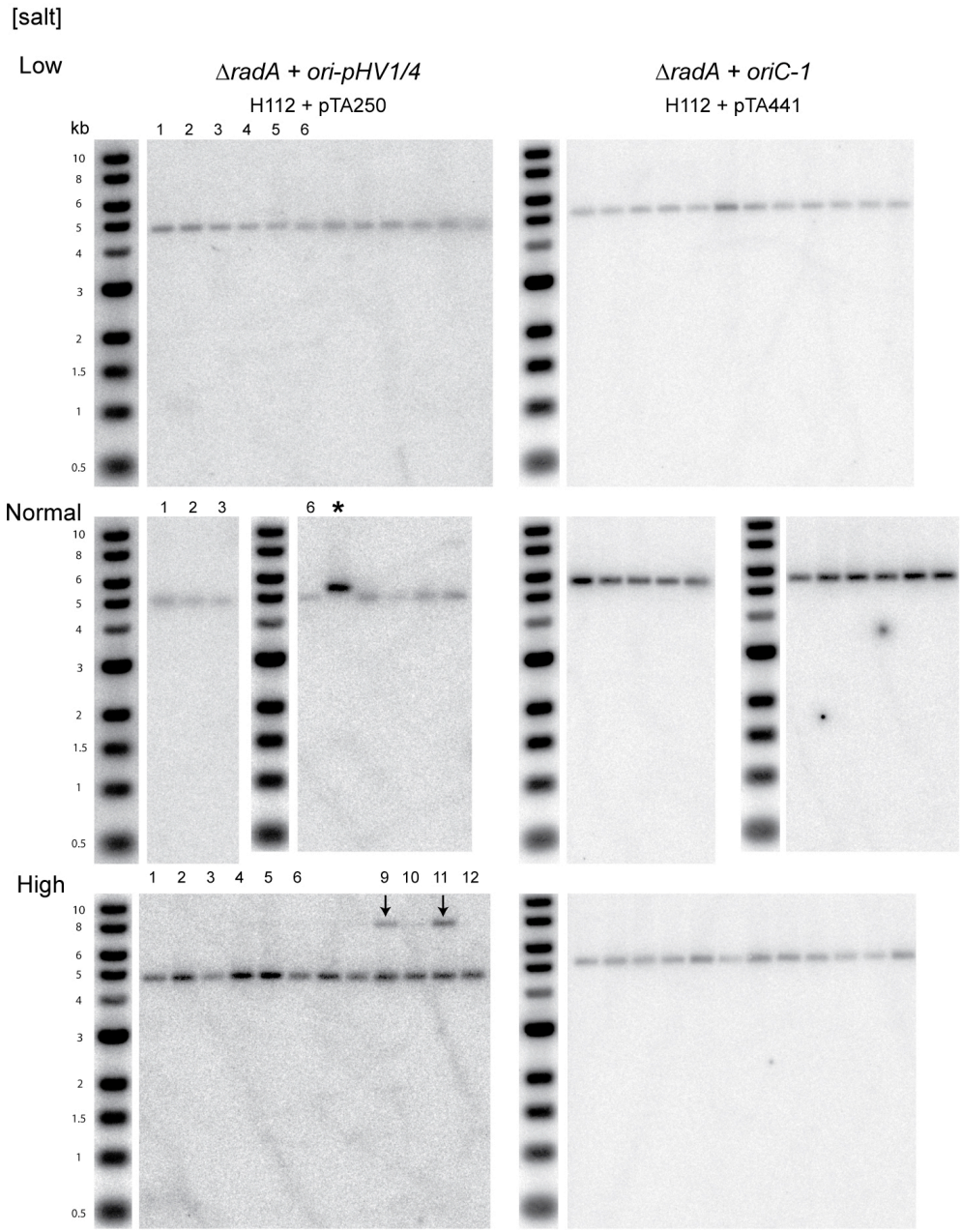


Figure 4.6: Further episomal plasmid maintenance assays in different salt conditions

Southern blot of crude DNA preparations from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*) or pTA441 (*oriC-1*). DNA was digested with *Eco*RI for pTA250 and *Bsp*EI for pTA441 transformants and the blot was probed with *pyrE2* (pTA52 *Nco*I/*Xba*I). Arrows mark the unidentified 8 kb band and the other bands represent linearised plasmid, pTA250 is 4.7 kb and pTA441 is 5.1 kb. The lane marked * is an incomplete digestion. Lane numbers refer to transformants examined in more detail in later experiments.

Panels A and B of Figure 4.7 show that the unidentified 8 kb band contains pBluescript II SK+ and *pyrE2* sequences and is not visible in uncut DNA. Uncut genomic DNA does not run in to gels efficiently. In the uncut lanes of panel C the *ori-pHV1/4* probe (derived from the cloned 8 kb fragment) hybridises with one high molecular weight band and the sample wells (not shown). Therefore, the unexpected 8 kb band sequences are not present on low molecular weight episomes. Panel C also confirms that *ori-pHV1/4* sequences are present in the unexpected 8 kb fragment.

In panel D of Figure 4.7 the restriction enzymes used all linearise the transforming plasmid (*ori-pHV1/4*, pTA250) except for *Cla*I. The 4.7 kb band represents the linearised *ori-pHV1/4* clone (pTA250). The presence of linearised plasmid bands means that the restriction digests were complete and therefore the obscure bands could not be the results of partial digestion leaving plasmid topoisomers. Since *Cla*I does not cut pTA250 this fragment is absent in *Cla*I lanes and replaced by a ~5.6 kb band that represents uncut supercoiled plasmid. The probe for Figure 4.7 panel D (pTA131) represents the vector for *ori-pHV1/4* subcloning used to generate pTA250. pTA131 is pBluescript II SK+ with the addition of the *pyrE2* marker. Using pTA131 as a probe for these various restriction digests of transformant DNA (Figure 4.7, panel D) shows that the larger hybridising DNA fragments vary in size but were not present in the negative control (transformant 12, no 8 kb band). This was further evidence that the fragments containing plasmid sequences were not episomal plasmid-related artefacts.

The presence of pBluescript II SK+, *pyrE2* and *ori-pHV1/4* sequences in non-episomal DNA suggested that pTA250 had integrated on to a *H. volcanii* chromosome. Integration should not be possible in this strain because RadA is the archaeal recombinase and Δ*radA* strains such as H112 are deficient in homologous recombination (Woods and Dyall-Smith, 1997). Despite this the band sizes in Figure 4.7, panel D represent what would be expected if pTA250 had integrated at the native *ori-pHV1/4* locus on pHV1. To confirm this, a diagnostic digest for this specific integration event was designed.

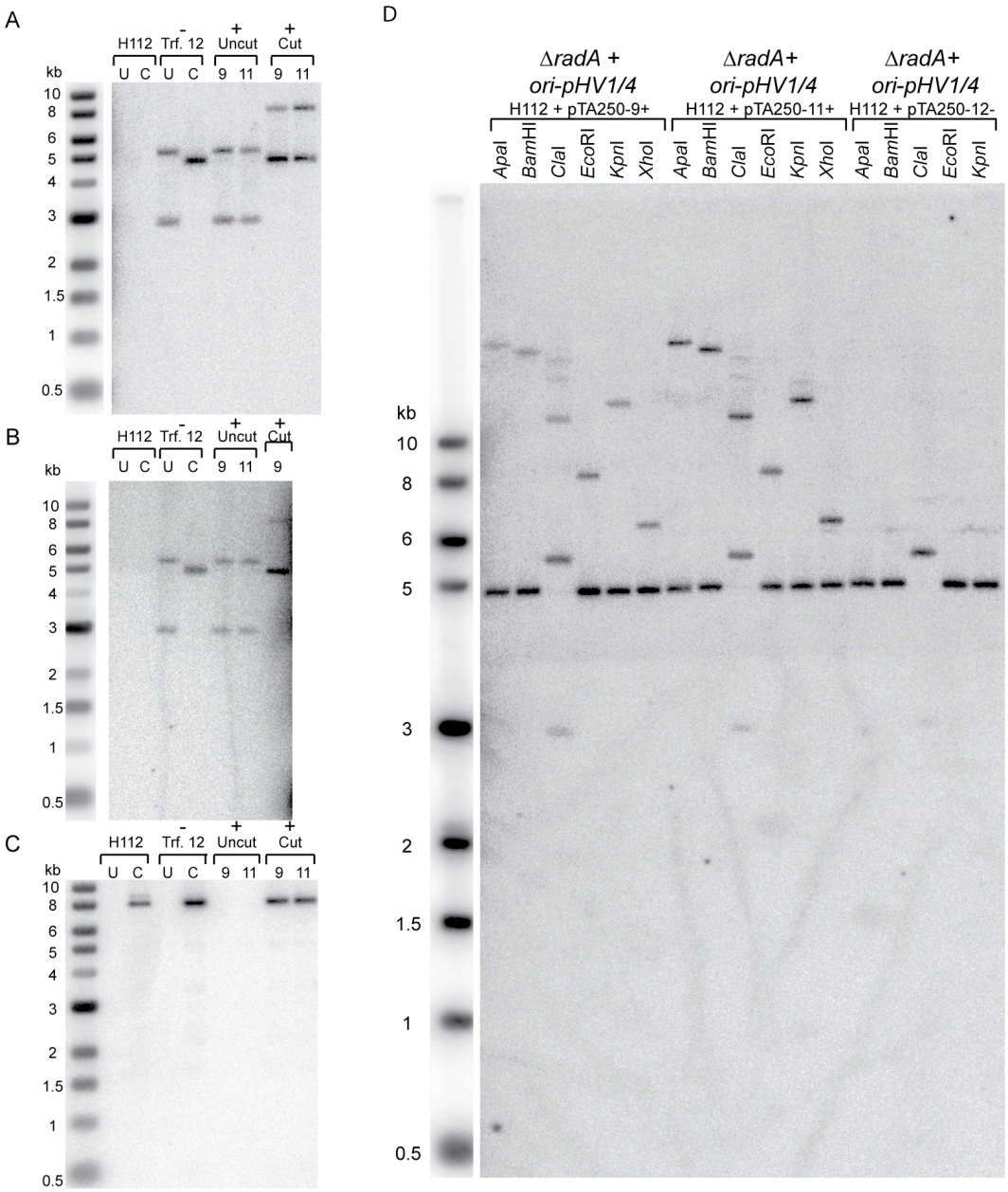


Figure 4.7: Analysis of transformants displaying an unidentified 8 kb band

Panels A-D show Southern blots of crude DNA preparations from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*), transformants (Trf.) 9, 11 and 12 from Figure 4.6. DNA was digested with *Eco*RI (C = cut) or run uncut (U= uncut).
+ = unidentified 8 kb band present. - = unidentified 8 kb band absent.
(A) Blot probed with *pyrE2* (pTA52 *Nco*I/*Xba*I).
(B) Blot probed with a pBluescript II SK+ specific probe (*Bgl*I/*Xmn*I).
(C) Blot probed with a fragment specific to an *ori-pHV1/4* fragment from the cloned unexpected 8 kb band (pTA894 *Eco*RI/*Xba*I).
(D) DNA was digested with the restriction enzyme shown and the blot was probed with the parental vector pTA131 (*pyrE2*$^{+}$, pBluescript II SK+).

Crude DNA from putative pTA250 integrants was digested with *Xho*I and the blot was probed with an *ori-pHV1/4* specific sequence derived from a subclone of ori-pHV1/4 (pTA252 *Nco*I/*Xba*I). The 6.1 and 7.3 kb bands that hybridised with this probe are diagnostic of a pTA250 integration event at the pHV1 *ori-pHV1/4* locus (Figure 4.8, panels A and B). The native and episomal *ori-pHV1/4* loci were also distinguishable. *H. volcanii* is polyploid (Breuert *et al.,* 2006) so this indicates that some replicons did not contain integrated pTA250 and these transformants are therefore meroploid. Using *pyrE2* as a probe for intact DNA displayed on a PFG from Δ*pyrE2* strains (Figure 4.8, panel C), a signal is visible for pHV1 and pHV4 when the *pyrE2* containing plasmid pTA250 has integrated. Therefore pTA250 integration occurs on pHV1 and pHV4 at regions of *ori-pHV1/4* homology and the shared region of these replicons containing this replication origin is at least 8.7 kb.

Integration of the *ori-pHV1/4* clone pTA250 explains the unexpected Southern blot banding patterns throughout this chapter (Figures 4.5-4.7). It also explains anomalous sequencing results obtained from the cloned 8 kb fragment in pTA894. This plasmid was constructed by cloning 8 kb fragments from a transformant that presented the unexpected 8 kb band in Southern hybridisations when *pyrE2* was used as a probe. Curiously pTA894 insert sequencing suggested that the cloned *ori-pHV1/4* fragment was ~4 kb when the insert fragments isolated from the library were 8 kb and this was confirmed by sequential sequencing across the entire insert. This cloning was an attempt to clone an integrated pBluescript II SK+-based plasmid in to pBluescript II SK+. With hindsight the sequencing results can be explained by sequencing primers annealing to various identical sequences within the same plasmid.

Since *radA*-independent recombination was so surprising, additional controls were performed using a *radA*$^+$ strain as a positive control and a Δ*ori-pHV1/4* strain as a negative control. 1 μg of pTA250 was used to transform H230 (*radA*$^+$, Δ*ori-pHV1/4*) and H26 (*radA*$^+$, Δ*pyrE2*) in normal salt conditions. These results are shown in Figure 4.8, panel C and Figure 4.9.

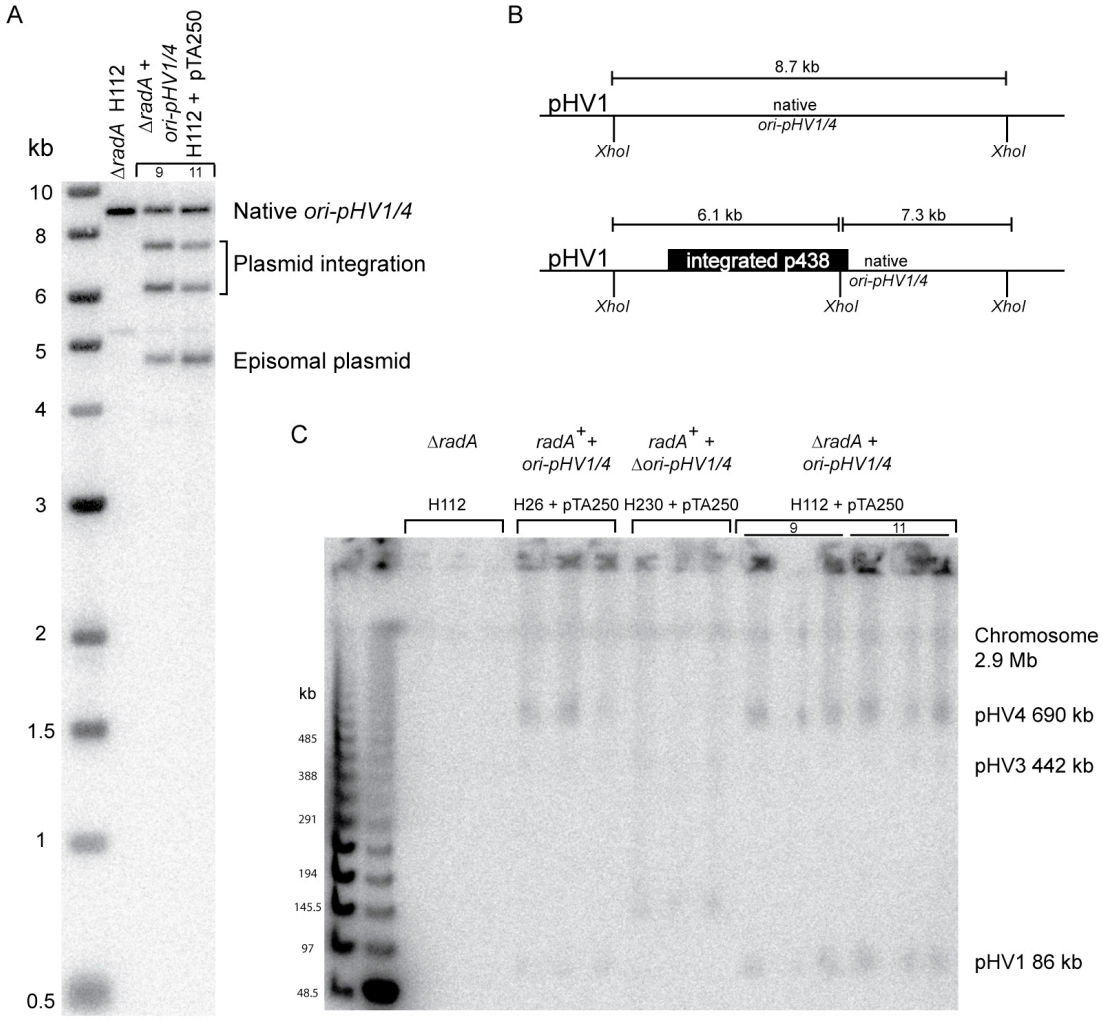


Figure 4.8: pTA250 integrated at the *ori-pHV1/4* locus in a Δ*radA* strain

(A) Southern blot of crude DNA preparations from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*), transformants 9 and 11 from Figure 4.6. DNA was digested with *Xho*I and the blot was probed with *ori-pHV1/4* (pTA252 *Nco*I/*Xba*I).

(B) Diagnostic digest and predicted fragments for pTA250 integration at the pHV1 *ori-pHV1/4* locus.

(C) Southern blot of intact *H. volcanii* DNA displayed on a PFG from H112 (Δ*radA*, Δ*pyrE2*), H26 (*radA*$^+$, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*), H230 (*radA*$^+$, Δ*ori-pHV1/4*) transformed with pTA250 and H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250. The blot was probed with *pyrE2* (pTA52 *Nco*I/*Xba*I).

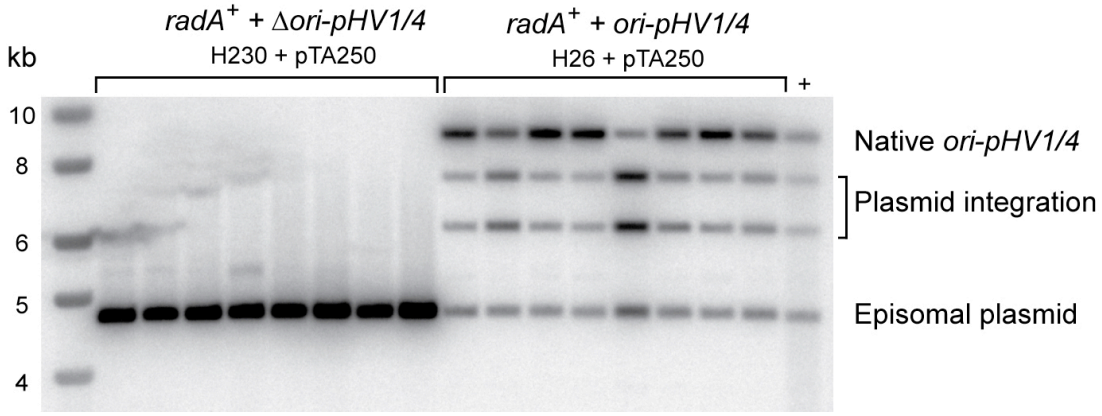


Figure 4.9: *radA*-independent recombination controls
Southern blot of crude DNA preparations from H230 (*radA*$^+$, Δ*ori-pHV1/4*) and H26 (*radA*$^+$, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*). DNA was digested with *Xho*I and the blot was probed with *ori-pHV1/4* (pTA252 *Nco*I/*Xba*I). The positive control (+) was transformant 9 (Figure 4.6 and 4.8).

Integration of pTA250 is likely in a *radA*$^+$ strain. This was reflected in Figure 4.9 with every transformant showing bands indicative of plasmid integration. Conversely, in the *ori-pHV1/4* deletion strain integration at the *ori-pHV1/4* locus is not possible because it has been deleted and only episomal plasmid was identifiable in Δ*ori-pHV1/4* transformants. This supports the observation that the integration process relies upon homology. These results were duplicated in the PFG shown in Figure 4.8, panel C where integration was determined by the presence of a *pyrE2* hybridisation signal for pHV1 and pHV4. This is visible for H26 (*radA*$^+$) lanes but is absent from H230 (Δ*ori-pHV1/4*) lanes.

The Δ*radA* deletion in H112 was verified by performing a PCR with primers that anneal to sequences within the deleted region of *radA*. When genomic DNA from the Δ*radA* strain (H112) was used as template DNA no amplification products were observed (Figure 4.10, panel B). Also shown in Figure 4.10 is the original confirmation of *radA* deletion in strain H112 using diagnostic *Mlu*I digestion and Southern hybridisation (performed by Hien-Ping Ngo). Both methods of confirming *radA* deletion demonstrate that the archaeal recombinase *radA* is completely deleted from strain H112. Therefore the observed integration of pTA250 occurs by *radA*-independent recombination.

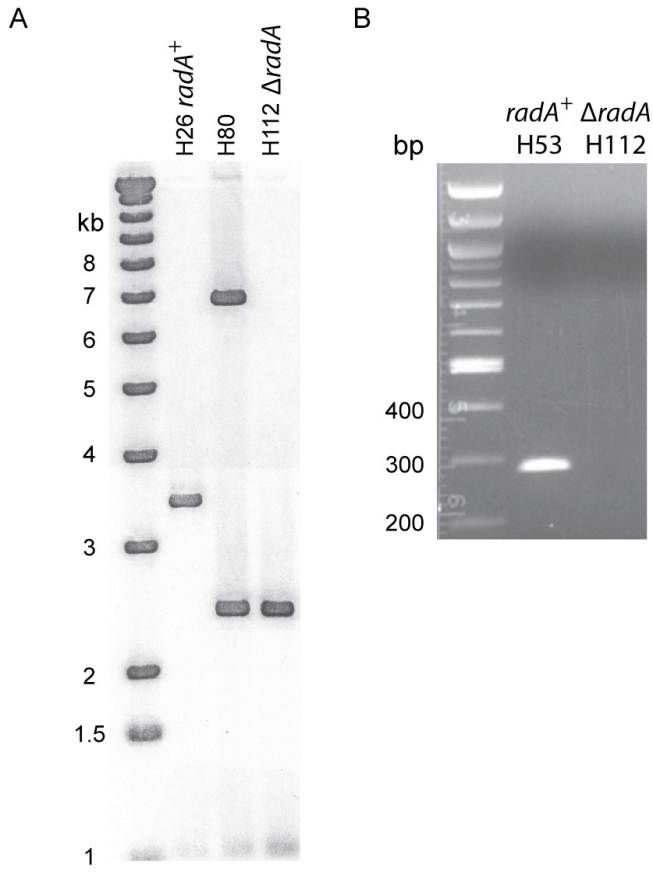


Figure 4.10: *radA* is deleted in H112

(A) *Mlu*I digest of genomic DNA from strains H26 (Δ*pyrE2*), H80 (Δ*pyrE2*, *radA*+::[Δ*radA pyrE2*+]) and H112 (Δ*pyrE2*, Δ*radA*), probed with DNA flanking *radA*. 1.2 and 3.3 kb bands indicate the native *radA* locus, the 7 kb band is indicative of deletion construct pop-in and the deleted *radA* locus is diagnosed by presence of 1.2 and 2.5 kb bands. Performed by Hien-Ping Ngo.

(B) PCR using primers radARTF and radARTR to amplify a 282 bp fragment of the *radA* gene. PCR was performed using Dynazyme with a 52°C melting temperature and a 30 second extension time. A "no template control" was also performed and no amplification products were visible.

4.3 *radA*-independent recombination

4.3.1 Stability of *radA*-independent recombinants

The consequences of pTA250 integration at the *ori-pHV1/4* locus are two-fold. Firstly, pHV1 and pHV4 will possess two copies of *ori-pHV1/4* separated by ~3.6 kb of foreign plasmid sequence. Secondly, the *pyrE2* marker will be present on both the episomal plasmids and the integrated copies. Therefore it should be possible to lose either the episomal or the integrated *pyrE2* genes by relieving selection for uracil. Curing of the episomal plasmid is more likely as previous results showed that pTA250 is only partially maintained without selection (Figure 3.11). Loss of the integrated plasmid is also possible. This would occur by intramolecular recombination which is the foundation of the pop-in pop-out strain construction methods used in this work.

Transformants 9-12 from Figure 4.6 were grown with selection in 5 ml high-salt Hv-Ca broth in an attempt to cure the plasmid so that integrants could be studied without episomal plasmid present. When the cultures reached exponential phase (OD_{650} 0.3-0.6) the cells were diluted 50-fold in to non-selective high-salt Hv-YPC broth. This was repeated twice with fresh high-salt Hv-YPC broth and before each dilution a crude DNA preparation was made from the exponential cells. This total DNA was digested with *Xho*I to diagnose integration status and plasmid loss.

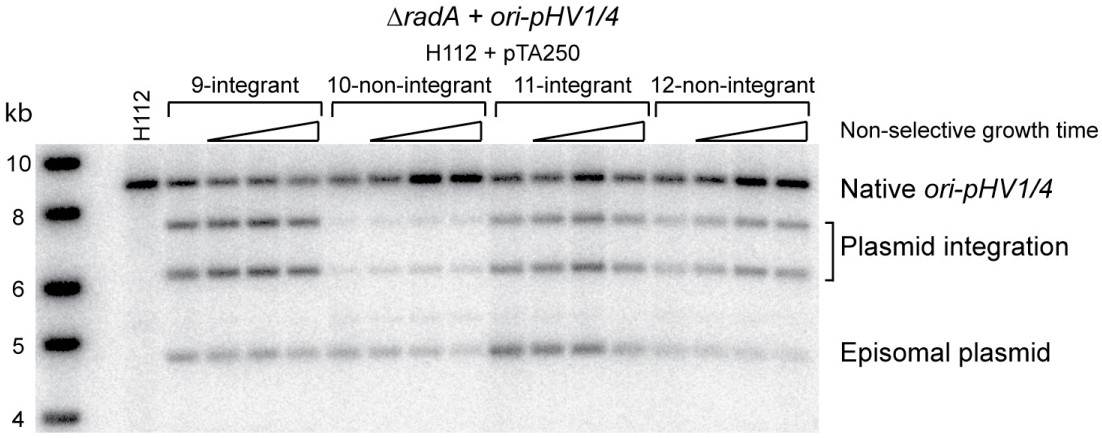


Figure 4.11: Stability of integrated and episomal pTA250-1
Southern blot of crude DNA preparations from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*) after non-selective growth (first lane for each sample) and with prolonged non-selective growth. DNA was digested with *Xho*I and the blot was probed with *ori-pHV1/4* (pTA252 *Nco*I/*Xba*I).

Analysis of the total DNA preparations presented two unexpected results. Transformants that were previously shown to be negative for the integration event (Figure 4.6, 10 and 12) now showed the 6.1 and 7.3 kb bands indicative of integration. The intensity of the plasmid integration bands for transformant 10 were low whereas the native *ori-pHV1/4* and episomal plasmid bands were of a similar intensity to the other samples. Weakly hybridising integration bands suggest that the amount of integration was less for this transformant. As the time grown without selection increased there was no corresponding loss in either episomal or integrated plasmid signals. This was surprising because in 3.11, pTA250 demonstrated only partial maintenance without selection and was lost at a rate of 16% per generation.

To assess whether conversion of non-integrant transformants to integrants was common, transformants which were negative for integration events in Figure 4.6 were examined again using the *Xho*I diagnostic digest.

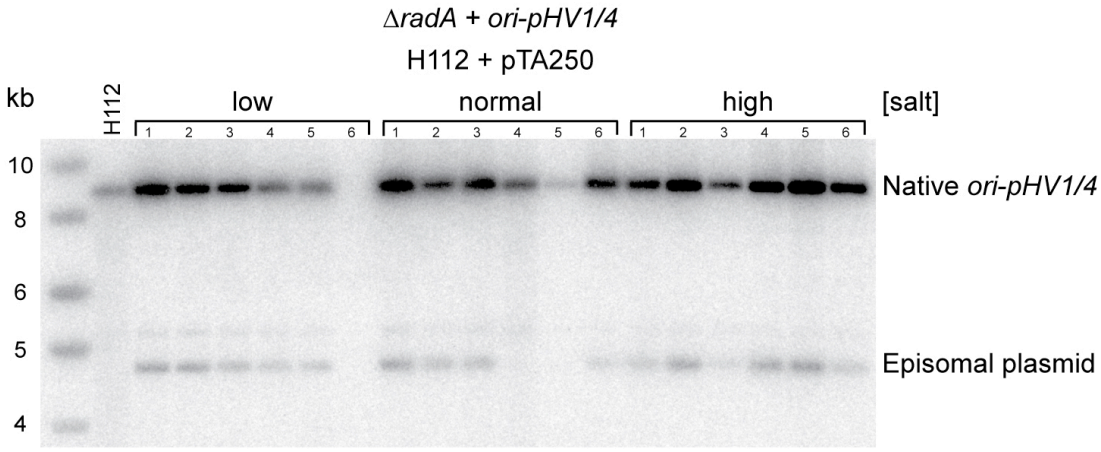


Figure 4.12: Non-integrants
Southern blot of crude DNA preparation from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*). Samples are transformants 1-6 from Figure 4.6 except for normal salt transformants 4 and 5 which were not shown. DNA was digested with *Xho*I and the blot was probed with *ori-pHV1/4* (pTA252 *Nco*I/*Xba*I).

None of these transformants presented the 6.1 and 7.3 kb bands which are diagnostic of integrated plasmid so the stability test was repeated using transformants 1 and 2 as the negative integration controls. High-salt *ori-pHV1/4* transformants 1, 2, 9 and 11 from Figure 4.6 were grown with selection in 5 ml high-salt Hv-Ca broth until exponential phase (OD_{650} ≈0.4-0.6), when they were diluted 250-fold in to fresh non-selective high-salt Hv-YPC broth. This was repeated twice with fresh high-salt Hv-YPC broth and before each dilution a crude DNA preparation was made. Before each

dilution cells were washed in 18% salt water and plated on high-salt Hv-YPC and Hv-Ca plates. 119 of these colonies were patched on to selective high-salt Hv-Ca plates to quantitatively measure plasmid retention during non-selective growth.

Table 4.2: Quantitative stability of integrated and episomal pTA250
Ability to grow on selective media after growth (0-3 serial dilutions) in non-selective conditions.

Transformant	Integrated pTA250	% cells retaining uracil marker after serial dilutions in non-selective media			
		0	1	2	3
1	-	87.4	78.2	79.8	74.8
2	-	34.5	21.8	14.3	20.2
9	+	100	100	100	97.5
11	+	99.2	97.5	95	99.2

Transformants with integrated pTA250 (9, 11) showed much higher maintenance of the ability to grow on media lacking uracil than transformant 2 which only contained episomal pTA250. This could confirm the hypothesis that episomal plasmid curing is more common than pop-out of an integrated plasmid. Transformant 1 was less able to grow on selective media after non-selective growth than integrants 9 and 11 but much more able than transformant 2. This suggested that transformant 1 had switched from being negative for integration to positive and this was confirmed by Southern blot analysis (data not shown). To test whether the reduction in the ability to maintain prototrophy for uracil of transformant 2 was due to plasmid curing, total DNA preparations from each round of serial growth were digested with *Xho*I and analysed as before (see Figure 4.13).

Figure 4.13 contradicts figure 4.12 because it shows transformant 2 with integrated pTA250. As was the case for transformant 10 (Figure 4.11), new integrant bands showed only a weak hybridisation signal indicating that across the population there had been fewer integration events than in other transformants. Transformant 11 was always positive for integration but in Figure 4.13 showed a much weaker native *ori-pHV1/4* locus signal than when it was previously analysed in figures 4.8 and 4.11. This points to integration having occurred on more copies of pHV1 and pHV4.

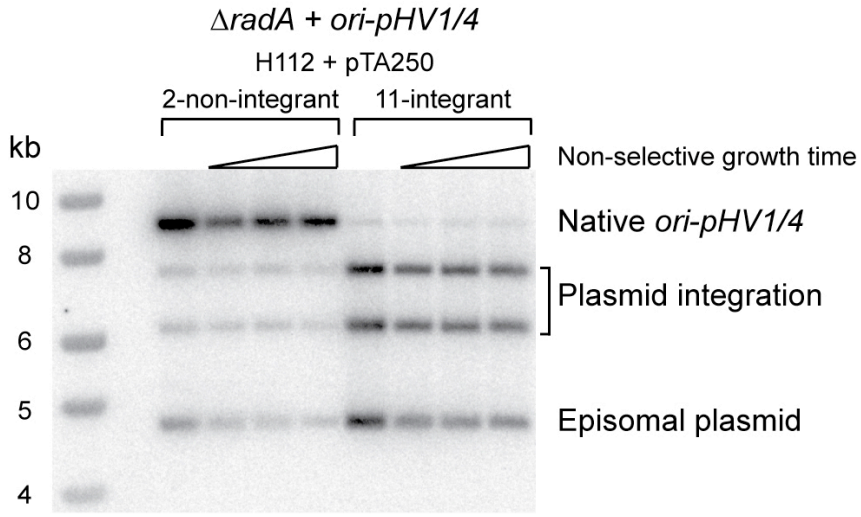


Figure 4.13: Stability of integrated and episomal pTA250-2
Southern blot of crude DNA preparations from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*) after non-selective growth (first lane for each sample) and with prolonged non-selective growth. DNA was digested with *Xho*I and the blot was probed with *ori-pHV1/4* (pTA252 *Nco*I/*Xba*I).

During the investigation of *radA*-independent *ori-pHV1/4* integration, transformants of interest were propagated by restreaking. Total DNA preparations were made fresh for every Southern blot because these crude preparations degrade quickly as a result of nuclease activity. The multiple examples of integrant status changes suggest that plasmid integration is a dynamic event which continues during growth and during room temperature storage of *H. volcanii* transformants. Since pTA250 contains the *ori-pHV1/4* replication origin it can propagate episomally. If non-selective growth causes pop-out of an integrated plasmid then a subsequent pop-in of episomal plasmid is likely. Conversely, if episomal plasmid is cured during non-selective growth it can be regenerated by pop-out of the integrated copy. This model is depicted in Figure 4.14 and explains why it was not possible to visualise episomal plasmid curing or pop-out of integrated plasmid over increasing amounts of non-selective growth time (Figures 4.11 and 4.13).

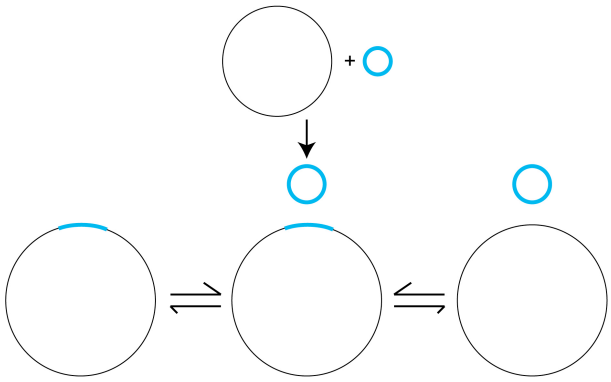


Figure 4.14: Model of dynamic integration
Blue represents transforming plasmid DNA, black represents a *H. volcanii* replicon. Due to continual episomal plasmid replication and plasmid integration, the equilibrium is always shifted towards the middle species of both episomal and integrated plasmid.

To test this more directly, plasmid integration was assessed in patched colonies that were three weeks old. Crude DNA preparations were made from four colonies that were patched on to Hv-Ca after the final round of non-selective growth for the stability experiment described in on page 152.

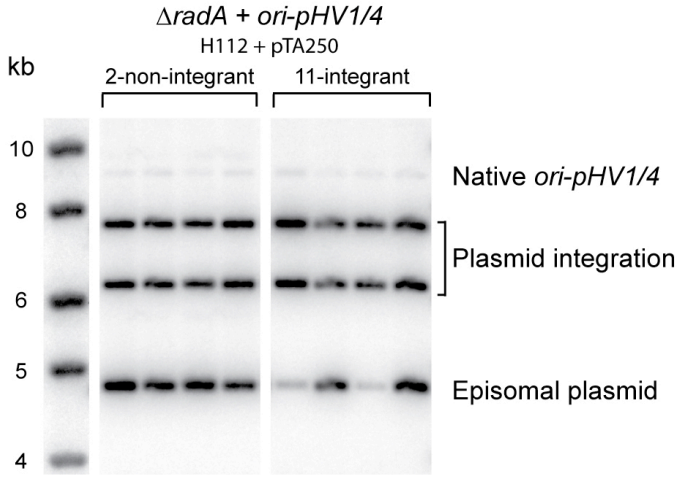


Figure 4.15: Stability of integrated and episomal pTA250 after prolonged storage
Southern blot of crude DNA preparations from individual colonies of H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*) after prolonged non-selective growth followed by selective growth. DNA was digested with *Xho*I and the blot was probed with *ori-pHV1/4* (pTA252 *Nco*I/*Xba*I).

In each of these colonies the native *ori-pHV1/4* signal was fainter than the plasmid integration bands. This indicated that integration was closer to occurring on every copy of pHV1 and pHV4 and the equilibrium of possible species was shifted towards integrated and episomal plasmid.

4.3.2 *radA*-independent recombination constraints

To begin to characterise constraints on the *radA*-independent recombination pathway, a high throughput assay was designed so that more transformants could be tested simultaneously. Since *radA*-independent recombination has only been observed for an *ori-pHV1/4* clone at high and low-salt concentrations it needs to be determined whether this mechanism can apply to different sequences and conditions.

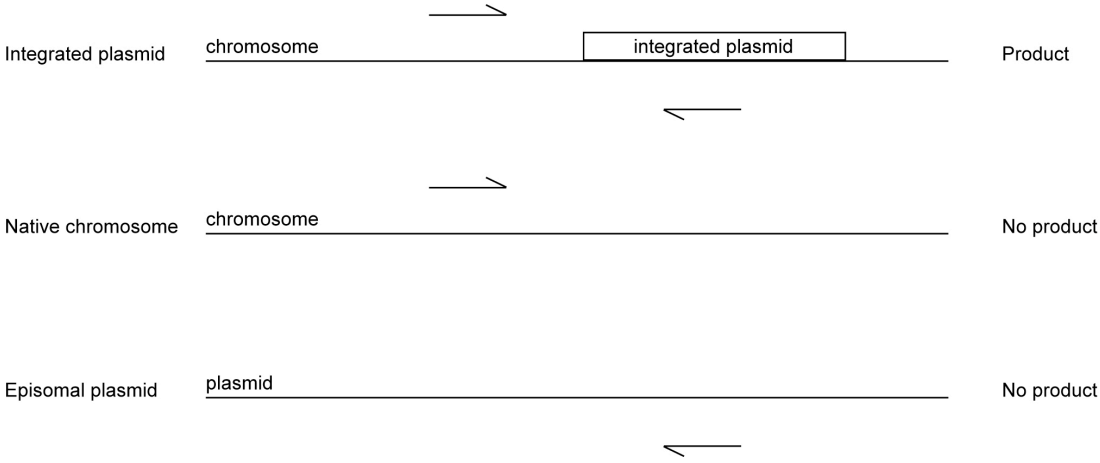


Figure 4.16: PCR assay for integrated plasmid
PCR strategy to test for plasmid integration using a chromosome-specific and plasmid-specific primer.

Using the strategy depicted in Figure 4.16, only an integrated plasmid would result in a PCR product. PCR was used to test additional *ori-pHV1/4* transformants at different salt concentrations. Fresh transformations of H112 (Δ*radA*, Δ*pyrE2*) with pTA250 were carried out in different salt conditions and 40 transformants were analysed for each salt concentration. Crude DNA preparations were made from patched transformants for the PCR template (as opposed to restreaking colonies and growing cultures in liquid broth to prepare crude DNA), to avoid excessive propagation that could alter integrant frequency estimates.

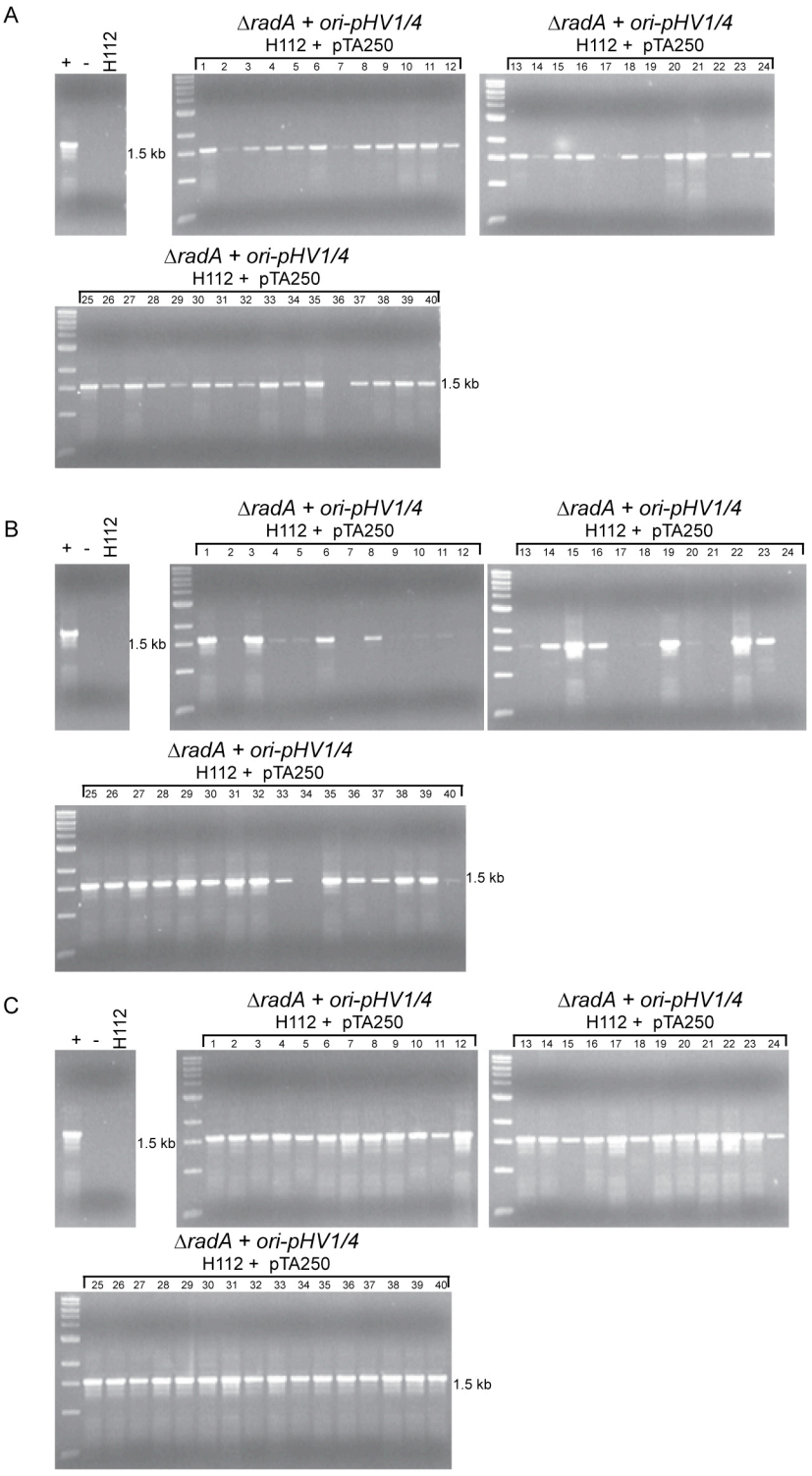


Figure 4.17: Integrant PCR assay for pTA250 H112 transformants

Ethidium bromide staining of PCR-based integrant assay products performed on crude DNA preparations from 40 pTA250 (*ori-pHV1/4*) H112 (Δ*radA*, Δ*pyrE2*) transformants using Chr-F and Plas-R primers. Transformations were performed at low (A), normal (B) and high (C) salt concentrations. A 1.5 kb PCR product is indicative of plasmid integration. The positive control (+) was a H26 (*radA*$^{+}$, Δ*pyrE2*) transformant DNA template and the negative control (-) was a non-integrant template.

All 40 high-salt transformants tested with the PCR-based assay showed pTA250 integration. At low and normal salt concentrations integration occurred at a frequency of 95% and 87.5% respectively. Variable PCR product concentrations suggested that integration occurred with variable efficiency. Based on the suggestion of a dynamic equilibrium for this process, the integration frequency is likely to be higher in low and high-salt transformants because they take longer to form colonies. The longer the generation time, the more likely integration is to have occurred. The PCR-based assay measure of integration frequency is significantly higher than the 17% figure derived from Southern blot analysis in Figure 4.6 so the DNA diagnostic digest and Southern blotting approach was used on the same transformants to see if the PCR results were corroborated. Transformants checked by Southern blotting were selected to cover the full range of PCR product band intensities shown in Figure 4.17.

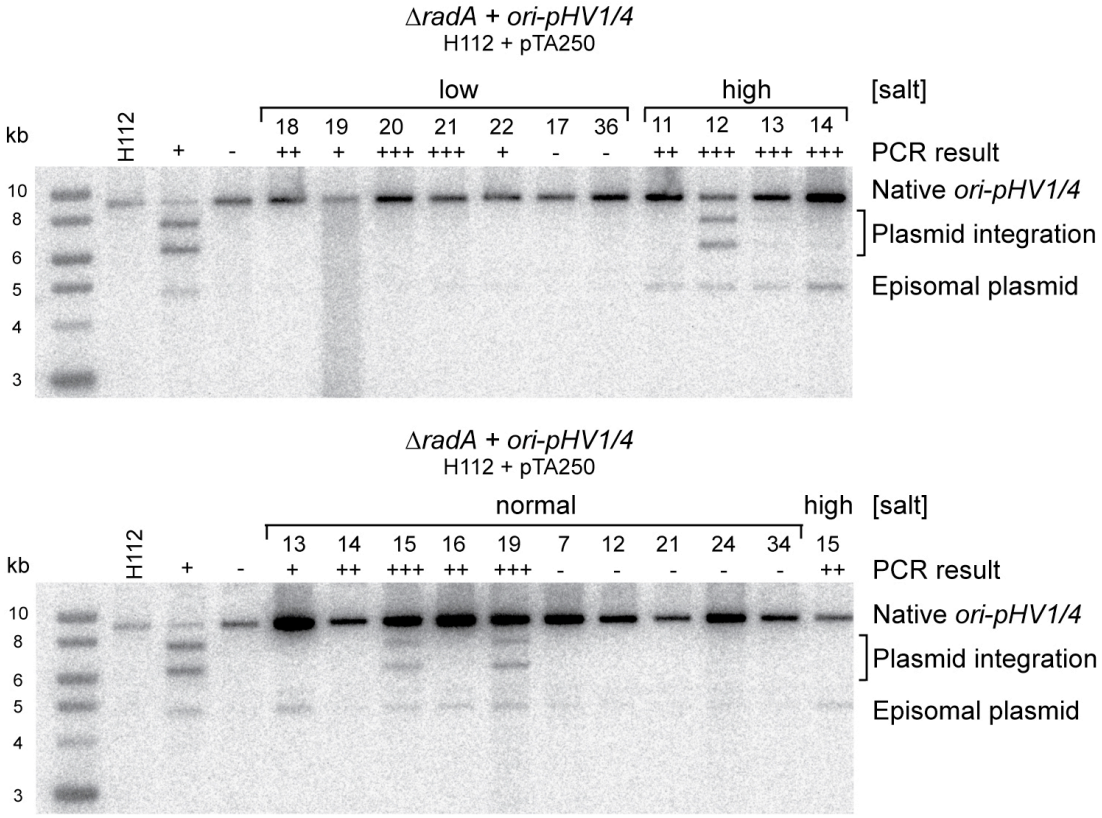


Figure 4.18: Southern and PCR integration assay comparison
Southern blot of crude DNA preparation from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA250 (*ori-pHV1/4*). DNA was digested with *Xho*I and the blot was probed with *ori-pHV1/4* (pTA252 *Nco*I/*Xba*I). PCR assay product intensities are indicated above each lane.

These two methods of checking for plasmid integration gave different results because PCR is inherently more sensitive than DNA-DNA hybridisation. The higher sensitivity of PCR meant that it gave false positives when compared to the Southern analysis in Figure 4.18 (for example transformants 18-22, low-salt). Southern blot analysis is quantitative once you have passed the minimum detection threshold and can be used to deduce the non-integrants vs. integrants equilibrium position of a population. The value of the PCR assay is that it provides a quick high-throughput route to test for presence or absence of integration events. The PCR assay showed that *radA*-independent integration is not dependent on a high-salt concentration.

The types of sequences that can integrate in a *radA*-independent manner needed to be characterised to determine the mechanism of the *radA*-independent recombination pathway. A variety of negative controls were used in the ARS assays described in Chapter 3. These included *leuB*, *bgaHv*, *oriC-X* and the intergenic region adjacent to *polX* cloned on a non-replicating plasmid (pTA131). These plasmids were used to transform H112 (Δ*radA*, Δ*pyrE2*) and produced no transformants, which means that these sequences do not function as replication origins in this assay. It also means that these sequences are incapable of integrating on to the chromosome because integrants would be able to grow on selective media. These experiments were therefore also controls for *radA*-independent recombination. In addition to these controls, previous studies of Δ*radA* strains have never observed homologous recombination (Woods and Dyall-Smith, 1997; Haldenby, 2007; T. Allers, unpublished observations).

oriC-1 and *oriC-2* transformants were analysed for plasmid integration to test whether *radA*-independent recombination was replication origin-specific. To further reduce the number of generations between plasmid transformation and DNA preparation, crude DNA was made directly from the original transformation plates using the "stab method" (2.2.7). This DNA was used as a template for the PCR-based assay and in Southern blot analysis.

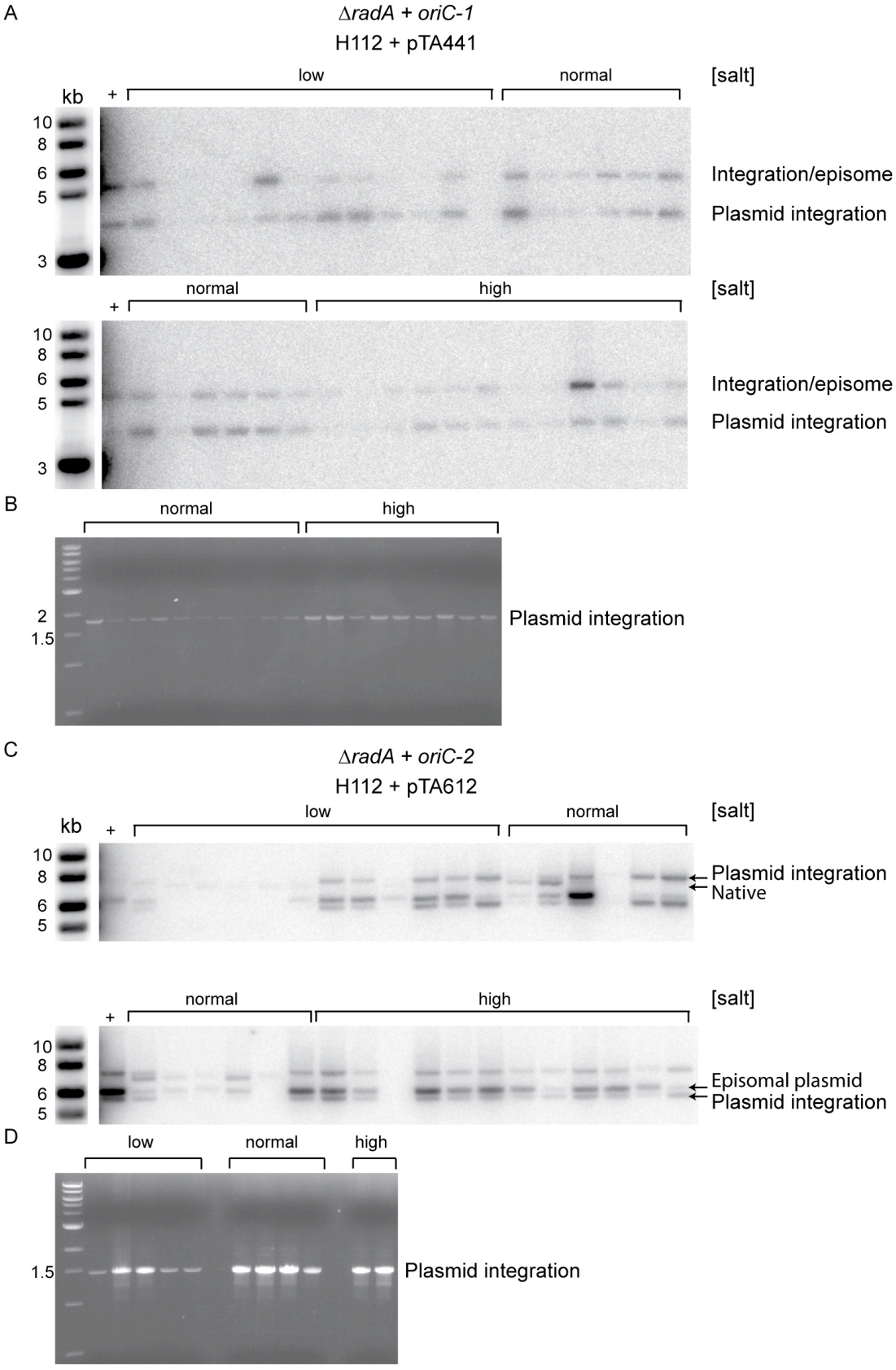


Figure 4.19: Integrated *oriC-1* and *oriC-2* ARS plasmids in a Δ*radA* background

Figure 4.19: **Integrated *oriC-1* and *oriC-2* ARS plasmids in a Δ*radA* background**
(A) Southern blot of crude DNA preparations from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA441 (*oriC-1*) in different salt conditions. DNA was digested with *Bsp*EI and the blot was probed with *oriC-1* (pTA441). The positive control was H26 (*radA*$^{+}$, Δ*pyrE2*) transformed with pTA441.
(B) Ethidium bromide staining of PCR-based integrant assay reaction products performed on crude DNA preparations derived from pTA441 (*oriC-1*) H112 (Δ*radA*, Δ*pyrE2*) transformants using PlasRC and *oriC-1* chrR2 primers. A 1.8 kb PCR product is indicative of plasmid integration.
(C) Southern blot of crude DNA preparations from H112 (Δ*radA*, Δ*pyrE2*) transformed with pTA612 (*oriC-2*) in different salt conditions. DNA was digested with *Bsp*EI and the blot was probed with *oriC-2* (pTA612). The positive control was H26 (*radA*$^{+}$, Δ*pyrE2*) transformed with pTA612.
(D) Ethidium bromide staining of PCR-based integrant assay reaction products performed on crude DNA preparations from pTA612 (*oriC-2*) H112 (Δ*radA*, Δ*pyrE2*) transformants using *oriC-2* chr-F and Plas-R primers. A 1.5 kb PCR product is indicative of plasmid integration.

Figure 4.19 shows PCR and Southern results which indicate that *oriC-1* and *oriC-2* based plasmids are capable of integrating on to the main chromosome in a Δ*radA* strain. Figure 4.19 does not present a contradiction with Figure 4.6 because the original *Bsp*EI digest could not discern between integrant and non-integrant transformants. As with *ori-pHV1/4*, *oriC-1* and *oriC-2* plasmids integrate at their native locus and in all salt conditions. The 5.9 kb native *oriC-1* band is not visible in panel A which could mean that all copies of the main chromosome have an integrated copy of pTA441.

The integration capability of *ori-pHV1/4*, *oriC-1* and *oriC-2* plasmids is evidence that *radA*-independent recombination is specific to replication origins. This could be linked to a property inherent to origin sequences. Alternatively perhaps only origin-based plasmids persist long enough and exist in a high enough copy number within cells for *radA*-independent recombination to occur. To distinguish between these possibilities a plasmid was constructed which contained two regions of *H. volcanii* homology, one of which was *ori-pHV1/4* to enable episomal replication. If *radA*-independent recombination was origin specific, integration should only occur at the *ori-pHV1/4* locus but if this pathway only requires time then integration could also occur at the second region of homology.

β-galactosidase was used as the second region of homology to enable quick screening of the possible transformation outcomes. The *H. volcanii β*-galactosidase gene (*bgaHv*) is not functional and does not hydrolyse X-gal to a blue precipitate. However, the *β-galactosidase* gene (*bgaH*) from the closely related species *Haloferax alicantei* is functional in *H. volcanii*. The strain used for this experiment contains a hybrid of the *bgaH* and *bgaHv* alleles, termed *bgaHa* (Delmas *et al*., submitted). An inactivating +2 frameshift mutation was generated at the *Kpn*I site by oligonucleotide insertion using PCR. Heteroalleles of *bgaHa* have been constructed with different inactivating mutations generated by oligonucleotide insertions (Haldenby, 2007). The *bgaHa-Bb* heteroallele was used to provide a second region of homology on *ori-pHV1/4*-based plasmids. Recombination between the mutations on the plasmid *bgaHa-Bb* allele and the chromosomal *bgaHa-Kp* allele (mutations situated at the *Bst*BI and *Kpn*I sites respectively) results in formation of the active *bgaHa* allele (see Figure 4.20).

The plasmids used are shown in Figure 4.20. pTA354 is a shuttle vector constructed by cloning *ori-pHV1/4* in to pTA131 (Norais *et al.,* 2007) and pTA151 is a non-replicating plasmid containing *bgaHa-Bb* (Haldenby, 2007). pTA641 was constructed by replacing the 784 bp *Nco*I/*Bst*XI fragment of the pTA354 shuttle vector with the functional *bgaHa* gene from pTA274 (*Nco*I/*Bst*XI) by Thorsten Allers. pTA1045 was made by replacing part of the *bgaHa* gene from pTA641 with the 774 bp *Eco*RI fragment of pTA151 so that it contained the inactive *bgaHa-Bb* allele. H607 (Δ*pyrE2*, Δ*trpA*, *bgaHa-Kp*, Δ*radA*::*trpA*$^{+}$) was transformed with 1 µg of each of these plasmids. Once colonies were visible they were sprayed with X-gal solution to assess recombination loci (see Table 4.3).

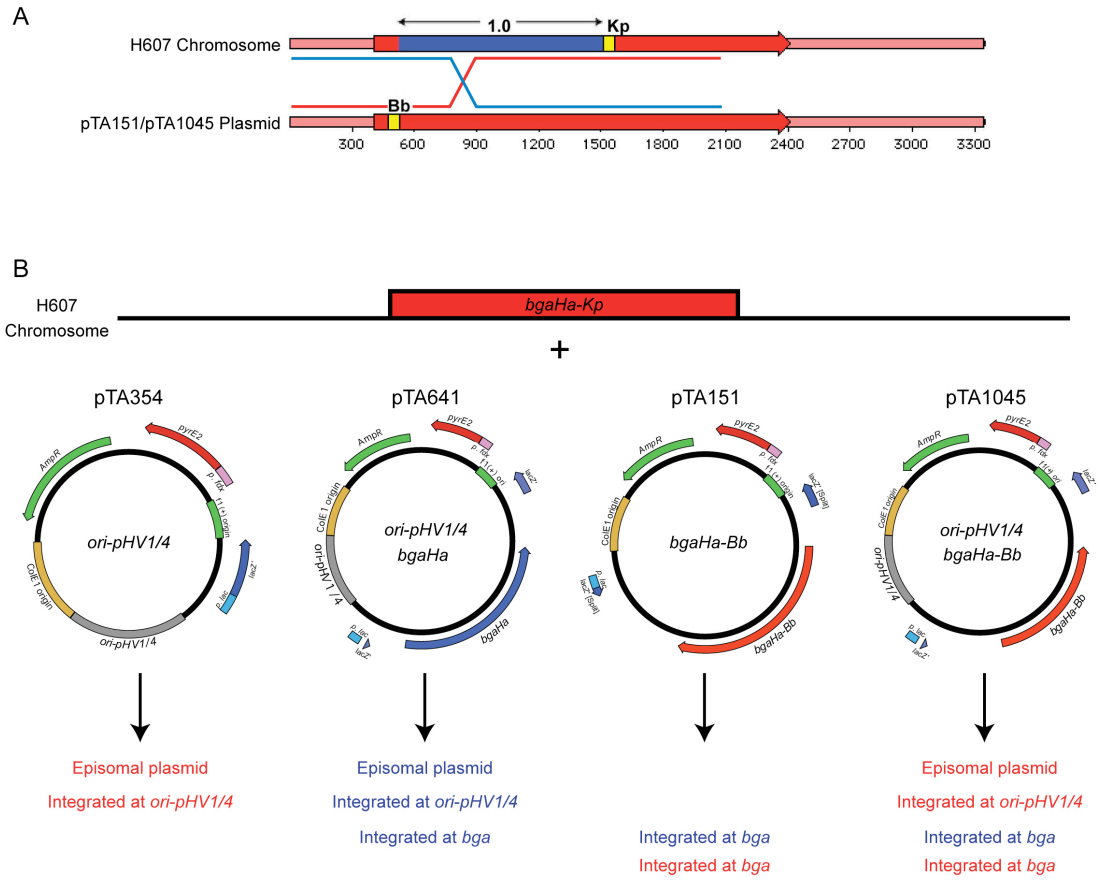


Figure 4.20: *radA*-independent recombination outcomes for *bga* and *ori-pHV1/4* plasmids

(A) *bga* heteroalleles of the host strain and pTA151/pTA1045. Yellow boxes represent the oligonucleotide insertions at unique restriction sites which generate an inactivating +2 frameshift mutation. The blue region represents the 1 kb of homology within which recombination can reconstitute a functional *bgaHa* allele. Potential crossover products are shown in blue or red according to the *bga* allele they contain. (B) Colours represent the possible outcomes of assaying H607 (Δ*pyrE2*, Δ*trpA*, *bgaHa-Kp*, Δ*radA*::*trpA*$^+$) transformants with X-gal. pTA151 and pTA1045 transformants can be red or blue depending upon the crossover location, see A.

Table 4.3: *bga* and *ori-pHV1/4* transformation results

Percentage of red/blue stained H607 (Δ*pyrE2*, Δ*trpA*, *bgaHa-Kp*, Δ*radA*::*trpA*$^+$) and H292 (see text for details) transformants after X-gal spraying. Data is from one experiment.

Strain	Transformant colour	Transforming plasmid			
		pTA354	pTA641	pTA151	pTA1045
H607	Red Blue	100% 0%	37% 63%	0 0	100% 0%
H292	Red Blue	100% 0%	0 100%	53% 47%	98% 2%

There were no transformants for pTA151 (*pyrE2*$^+$, *bgaHa-Bb*) because this is a non-replicating plasmid control. This is further confirmation that *radA*-independent recombination does not normally occur at the *bga* locus. Plasmid pTA1045 (*pyrE2*$^+$, *ori-pHV1/*4, *bgaHa-Bb*) contains two possible sites for recombination. The lack of blue colonies suggested that no plasmid integration occurred at the *bga* locus. However, though the pTA354 (*pyrE2*$^+$, *ori-pHV1/4*) control transformation matched the predicted outcome by producing 100% red colonies; the pTA641 (*pyrE2*$^+$, *ori-pHV1/4*, *bgaHa*) control presented unexpected red transformants. This could occur by gene conversion of the *bgaHa* allele on the plasmid by the chromosomal *bgaHa-Kp* allele. However, the expected frequency of gene conversion would be lower than 37% (Haldenby, 2007).

Based on previous observations that there are problems with β-galactosidase expression in Δ*radA* strains (T. Allers, unpublished observations) this experiment was also performed in the strain H292 which is isogenic with H607 except for being *radA*$^+$. Transformation of H292 using the control plasmids gave the expected results. 100% of pTA354 transformants were red, 100% of pTA641 transformants were blue while pTA151 transformants showed the expected 50:50 mix of red and blue colonies. pTA1045, which can propagate episomally or integrate at either the *bga* or the *ori-pHV1/4* locus, gave a ~2% frequency of blue transformants indicating plasmid integration at *bga*.

The H292 pTA1045 transformation results (~2% blue) demonstrate that in a *radA*$^+$ strain, recombination at the *ori-pHV1/4* locus dominates recombination at the *bga* locus. Blue colonies were recovered from *bgaHa* transformants of a Δ*radA* strain with a 63% frequency. Therefore, it should be possible to rule out *radA*-independent recombination occurring at *bga* by expanding this assay to ensure there is enough statistical power to detect the potentially low number of blue colonies from pTA1045 (*pyrE2*$^+$, *ori-pHV1/*4, *bgaHa-Bb*) transformants.

4.4 Discussion

This chapter began by exploring the hypothesis that *H. volcanii* origins of replication were conditional. There are many possibilities for variables that could influence origin activity. Environmental salt concentration was chosen based on the predicted Orc pI variation and the wide salt tolerance of *H. volcanii*. An ARS assay performed with five *H. volcanii* replication origins in three different salt concentrations was used to indirectly test origin activity. Unfortunately the replication origin transformation efficiencies and colony formation times were quite variable and difficult to quantify. The most consistent result was the slow growth of *ori-pHV3* transformant colonies at low and high-salt concentrations. This could suggest that *ori-pHV3* is only active within a narrow range of salt concentrations. Of all the origin-associated *orc* genes, the *ori-pHV3*-associated *orc6* encodes the Orc protein with the highest theoretical pI (4.98). Therefore a possible mechanism for the narrow tolerance of *ori-pHV3* could be that the associated Orc6 can not bind to *ori-pHV3* repeats efficiently at non-optimal salt concentrations. However *H. volcanii* has multiple *orc* genes. Origins are not necessarily bound exclusively by their associated Orc. Distinct repeats at *H. volcanii* replication origins suggest they might be bound by different proteins but there have been no studies of Cdc6/Orc1-origin binding in *H. volcanii* so the binding profile of each Cdc6/Orc1is unknown. Cloned origin ARS plasmids do not replicate *in vivo* origin activity because they are not situated in their chromosomal context (see 3.10). For example, DNA topology will differ on an artificial plasmid compared to a larger chromosome and this might result in differences in origin activity. Therefore the ARS assay alone is not the best method to test for differential regulation of origins. A combination of *orc* expression analysis and Orc binding specificity studies would be more suited to addressing this question.

Instead of pursuing conditional origins, this chapter focused on the chance discovery of *radA*-independent recombination. In the early days of transformation experiments Struhl *et al*. described four possible transformant outcomes:

1) Autonomous replication of the transforming DNA.
2) Integration of the transforming DNA by homologous recombination.
3) Integration via illegitimate (non-homologous) recombination.
4) Miscellaneous transformation. (Struhl *et al.,* 1979)

ARS assays are based on finding category one transformants. The justification for performing the ARS assays in this work in a strain deficient in homologous recombination was to eliminate category two transformants. Nevertheless, this study established that a *radA*-independent recombination mechanism exists in *H. volcanii*. Evidence presented here suggests this pathway is replication origin-specific but has no direct relationship with salt-stress or a particular replicon. In each example of *radA*-independent recombination, the transforming plasmid integrated at its native locus. The Southern blots performed can not distinguish between a single integration event or a tandem array of integrated plasmids. Discerning between these possibilities will be important for further characterisation of this process. Integration of ARS plasmids at other loci would result in additional unidentified bands in the diagnostic Southern blots and these were not observed. Therefore *radA*-independent recombination is likely to involve sequence homology. The inability to recover integrant strains with no episomal plasmid suggests that this process is dynamic and that regeneration of episomes occurs by *radA*-independent pop-out of the integrated copy. The dynamics of this process are interesting and a quantitative assay such as real-time PCR would be of value in determining the rate and efficiency of this pathway. In Chapter 3 it was notable that *oriC-2*, *ori-pHV3* and *ori-pHV4-2* ARS plasmids were highly unstable (Figure 3.21). Perhaps for these plasmids the cloned origins were not efficient enough to keep the equilibrium position shifted towards the middle species shown in Figure 4.14, they could therefore be lost without selection.

This work has begun to address the key questions about the characteristics of *radA*-independent recombination but the mechanism involved has not been tackled directly. The origin-specificity of *radA*-independent recombination is a crucial question to answer in order to narrow down potential mechanisms. It should be possible to test whether the apparent origin-specificity is an artefact of the persistence of autonomously replicating plasmids in a cell. To determine if plasmid persistence explained the origin-specificity of *radA*-independent recombination a useful variation of the assay described in Figure 4.20 would be to use a plasmid based on a replication origin with no homology in the *H. volcanii* genome. Since *ori-pHV2* plasmids do not function in Δ*radA* strains (Woods and Dyall-Smith, 1997), a solution would be to transform a Δ*radA* strain deleted for a replication origin (e.g. *ori-pHV1/4*) with a plasmid containing that origin and a different region of homology. This plasmid would be capable of autonomous replication and so would

be maintained episomally long enough for potential integration at the second region of homology.

In ARS assays the cloned origin on the episomal plasmid is identical to the equivalent native chromosomal origin. Replisomes can be blocked by a variety of DNA lesions and replication can then be restarted using template switching to bypass the block. Normally this would occur by switching to the complementary DNA strand or another copy of the replicon in the case of polyploids. However in ARS transformants template switching can occur between the artificial episomal plasmid and chromosomes. Figure 4.21 depicts a model for origin-specific recombination based on a replication fork stall.

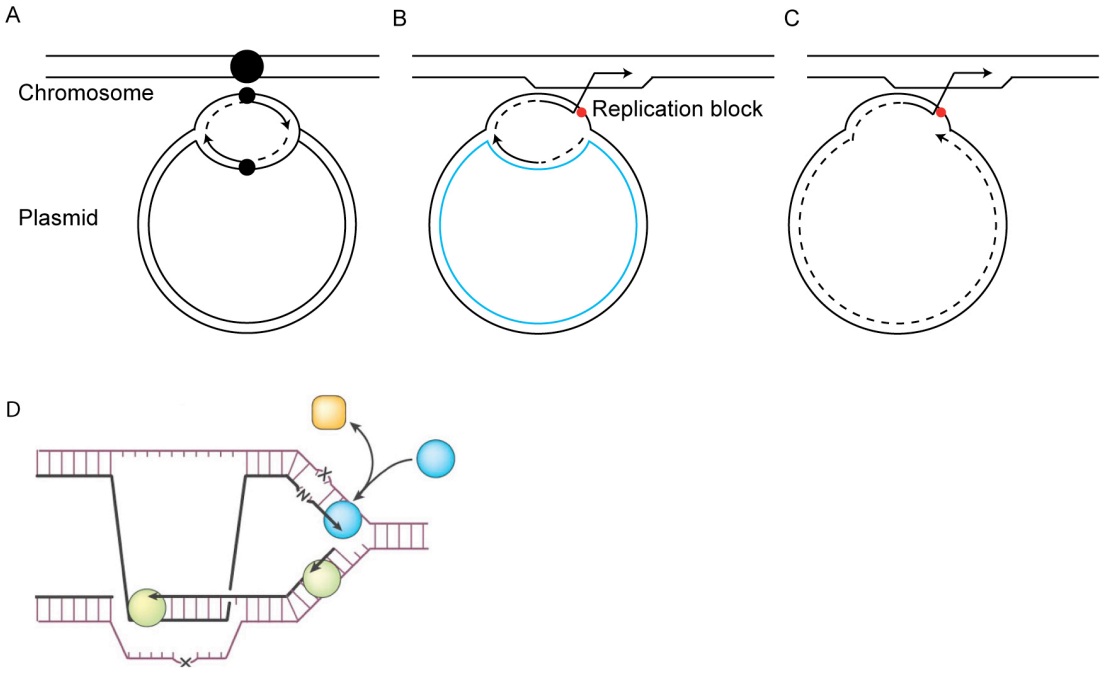


Figure 4.21: Model for origin-specific *radA*-independent recombination
(A) Episomal plasmid containing a chromosomal replication origin (black circle) which replicates using bidirectional DNA replication. Discontinuous lagging strand synthesis is represented by the dashed line.
(B) A leading strand replication block occurs at the bidirectional plasmid replication fork. Template switching enables replication to continue by priming from homologous chromosomal DNA. The blue strand completes normal replication.
(C) Lagging strand DNA synthesis continues for the unblocked DNA polymerase and resolution results in integrated plasmid.
(D) Detailed view of template switching (bottom strand). Green replicative polymerase bypasses damaged DNA by switching to a homologous template. From Hoeijmakers, 2001.

Replication blocks could be relatively common on ARS plasmids because the origins are not in their chromosomal context. Plasmid replication forks could stall because the *H. volcanii* replication machinery is replicating foreign DNA that differs in base content and topology. This model is not limited to replication blocks on the leading strand. However, strand invasion and template switching is less likely for discontinuously synthesised Okasaki fragments due to their length. The unidirectional replication shown in Figure 4.21, panel C, is similar to the rolling-circle mode of replication observed in bacterial and archaeal plasmids (reviewed in Khan, 2005). Since the chromosomes of *H. volcanii* are circular, this model could also apply to replication blocks on the chromosome. However this is less feasible due to the size disparities and in the case of the main chromosome, the presence of multiple replication origins. While this model can explain the origin-specificity of *radA*-independent recombination, the strand invasion and template switching step relies upon pairing homologous sequences. Since this occurs in the absence of RadA, an alternative recombinase may exist in *H. volcanii*. Plausible candidates include the two site-specific recombinase XerD homologues or the multiple putative phage integrases within the *H. volcanii* genome. Future work to determine the *radA*-independent recombination mechanism could involve deleting these genes and assaying for origin-based plasmid integration in a Δ*radA* strain. Another possibility is that replication occurs in distinct replication factory domains and homologous sequences pair without the need for a recombinase protein.

Whatever the mechanism is for *radA*-independent recombination, its existence has implications for the study of replication origins in Chapter 3. The ARS assay relies on the absence of false positives from recombination in order to determine origin activity. The possibility of *radA*-independent recombination highlights the importance of using *in vivo* tests of origin activity to confirm replication origins. However, if *radA*-independent recombination proves to be origin-specific then positive ARS assay results due to integration will not be false positives but will demonstrate another aspect of origin sequences. One of the controls for the ARS assays in Chapter 3 was a check for episomal plasmid maintenance. This was achieved by digesting total transformant DNA with an enzyme selected to linearise the transforming plasmid and hybridising the gel blot to a pBluescript II SK+ probe to identify transforming DNA (Figure 3.17). Each plasmid linearised to the correct size, indicated episomal maintenance, but the *oriC-2* transformant DNA also

presented a band >10 kb. With the benefit of hindsight from this chapter, the extra band can be identified as a 12.9 kb fragment that is indicative of pTA612 (*pyrE2*$^{+}$, *oriC-2*) integration on to the chromosome at the *oriC-2* locus. The other four transforming plasmids did not integrate, although there may be a faint 8 kb band indicative of integration for the *ori-pHV1/4* transformant DNA lane. Integration might not be evident for the other plasmids because of the short time span between transformation and isolation of DNA. It is possible that there is more selective pressure towards integration for *oriC-2*-based plasmids. *oriC-2* is low efficiency ARS because *oriC-2* transformants present small colonies and are have a low copy number (Table 3.5) Cells that have integrated the transforming plasmid are no longer dependent on the inefficient episomal *oriC-2* to replicate the *pyrE2* marker and enable growth.

RecA-independent recombination occurs in *E. coli* (Lovett *et al.,* 2002). The mechanism is unknown but has been shown to be DnaK-dependent (Goldfless *et al.,* 2006). RecA-independent recombination is limited by single-stranded exonucleases which suggests that ssDNA is involved (Dutra *et al.,* 2007). These authors speculated that RecA-independent recombination is targeted to replication forks and may contribute to genome evolution. Some features of RecA-independent recombination show similarities with the data presented here. Studies in *E. coli* may inform the future characterisation of *radA*-independent recombination such as an investigation of a link with the *H. volcanii dnaK* homologue.

Chapter 5: Characterisation of *RAD25*-like genes

5.1 Introduction

The region surrounding *oriC-2* contains three genes with homology to *S. cerevisiae RAD25* and a fourth *RAD25-like* gene is present on pHV4. These genes have been annotated as *RAD25-like helicases* based on their homology to the eukaryotic *RAD25* gene which is also known as *XPB* or *SSL2* (*S. cerevisiae*). *RAD25* is involved in DNA repair and also has a role in transcription.

To maintain genome integrity, multiple mechanisms to correct environmentally-induced DNA damage and DNA replication errors have evolved. These include direct damage reversal, base excision repair (BER), nucleotide excision repair (NER), mismatch repair and recombinogenic repair (Aravind and Koonin, 1999). Different repair pathways have different target lesions but some, like NER, can repair a broad spectrum of DNA damage. NER is universal and the biochemical principles are conserved across Bacteria and Eukarya. In NER, DNA lesions are recognised by their distortion of the DNA helix. Incisions are made either side of the DNA damage and the oligonucleotide fragment containing the damage is then excised. The final step is repair of the gap and ligation (Friedburg, 2006). Despite mechanistic conservation across the domains, the proteins involved are unrelated.

Table 5.1: NER factors in prokaryotes and eukaryotes

Prokaryotes		**Eukaryotes**	
Name	**Biochemical activity**	**Name**	**Biochemical activity**
UvrA	ATPase, DNA binding	NEF1: Rad1, Rad10, Rad14	DNA endonuclease, DNA damage recognition
UvrB	Helicase	NEF2: Rad4, Rad23	DNA damage binding Tethering of NEF1 with NEF3
UvrC	Nuclease	NEF3: Rad2, Rad3, Rad25, SSL1, TFB1, TFB2, TFB3	DNA endonuclease, DNA helicase
UvrD	Helicase	NEF4: Rad7, Rad16	DNA dependent ATPase, DNA damage recognition
		RPA: p69, p36, p13	DNA damage recognition

NEF = nucleotide excision repair factor protein assemblies.

Adapted from Prakash and Prakash, 2000.

The DNA repair genes of *S. cerevisiae* fall in to three epistasis groups: *RAD3*-excision repair, *RAD6*-post-replication repair and *RAD52*-recombinogenic repair. *RAD25* is in the *RAD3* epistasis group. *RAD25* was independently isolated as the *S. cerevisiae* homologue of the *H. sapiens* nucleotide excision repair gene *ERCC-3* (Park *et al.,* 1992) and as a suppressor of engineered stem-loop structures in mRNA where it was named SSL2 (Gulyas and Donahue, 1992). *ERCC-3* is a DNA helicase and defective versions result in the disease xeroderma pigmentosum (XP). XP is characterised by sunlight-induced hyperpigmentation of the skin and a predisposition to skin cancer (Lehmann, 2003). Hypersensitivity to ultraviolet (UV) radiation is indicative of an NER disorder because the induced 6-4 photoproducts and cyclobutane pyrimidine dimers are the most common DNA lesions repaired by NER. There are eight complementation groups for XP. Human *ERCC-3* complements the UV sensitivity in *ERCC-3* deficient rodent cell lines from complementation group XP-B (Weeda *et al.,* 1990a; Weeda *et al.,* 1990b), therefore this gene is also known as *XPB*. *S. cerevisiae RAD25* is the functional homologue of *XPB*/*ERCC-3*. *RAD25* nomenclature will be used in this work.

S. cerevisiae Rad25 contains the seven helicase motifs characteristic of superfamily II helicases, a nuclear localisation signal and a DNA binding domain (Park *et al.,* 1992). Rad25 has ssDNA-dependent ATPase and DNA helicase activity (3′-5′) (Guzder *et al.,* 1994), but unlike most *S. cerevisiae* NER factors, *RAD25* is essential. Specifically the ATPase and helicase activity is crucial for viability, suggesting that *RAD25* has another role during normal cell growth (Park *et al.,* 1992). *RAD3* (*XPD*) is also essential but Rad3 ATPase/helicase activity is not (Naumovski and Friedberg, 1983; Sung *et al.,* 1988). The essential role of *RAD25* was resolved when it was discovered to be a subunit of the basic transcription factor for RNA polymerase II, TFIIH (Schaeffer *et al.,* 1993; Guzder *et al.,* 1994; Wang *et al.,* 1994). RAD25 therefore has a dual role in transcription and repair. *RAD3* is also part of the TFIIH complex (Feaver *et al.,* 1993).

Rad25 is essential because TFIIH requires Rad25 to open DNA during transcription initiation (Coin *et al.,* 1999; Douziech *et al.,* 2000). Rad25 can correct the transcriptional defect of *rad25-ts* conditional lethal mutants (Qiu *et al.,* 1993; Guzder *et al.,* 1994). Studies in *S. pombe* showed that it is the ATPase activity of Rad25 which drives promoter opening. Helicase activity is required to allow escape of RNA

polymerase II from promoters so that it can enter elongation mode (Lin *et al.,* 2005). In NER the ATPase activity of Rad25 and the helicase activity of Rad3 is required for local unwinding prior to the incision step (Sung *et al.,* 1996; Evans *et al.,* 1997; Coin *et al.,* 2007). One model for NER is that Rad25 wraps around DNA to cause local melting which enables the helicase activity of Rad3 to create an unwound DNA bubble structure. This can then be cleaved by Rad2 (Xpg) (3′) and Rad1 (Xpf) (5′) and the DNA between this dual incision is excised (Coin *et al.,* 2007).

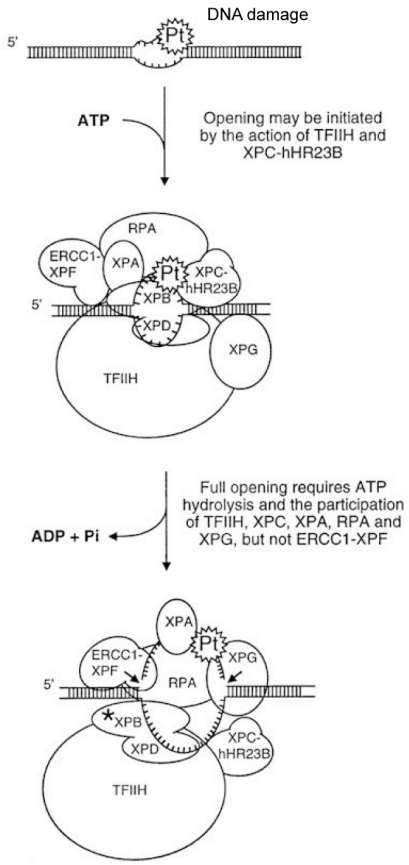


Figure 5.1: Model of eukaryotic NER
Open complex formation involving RAD helicase components of TFIIH followed by dual incision around the DNA damage by other NER proteins. From Evans *et al.,* 1997.

Archaeal DNA repair and transcription processes are more similar to those found in eukaryotes than bacteria. Many archaeal genomes contain homologues of eukaryotic NER genes (*RAD1*/*XPF*, *RAD2*/*XPG*, *RAD3*/*XPD* and *RAD25*/*XPB*) (Kelman and White, 2005). Mesophilic methanogens and halophiles also have homologues of the bacterial NER genes *UvrA*, *UvrB* and *UvrC*. This is likely due to recent lateral gene transfer from bacteria (White, 2003). The study of DNA repair in archaea lags behind that of eukaryotes. NER has only been demonstrated in Archaea with the *Uvr*

homologues (Ogrunc *et al.,* 1998; McCready and Marcello, 2003). No components of TFIIH except for *RAD3* and *RAD25* have been identified in archaea (Thomm, 2007). The crystal structure of *A. fulgidus* Rad25 suggests that it recognises DNA damage-induced distortions rather than specific lesions (Fan *et al.,* 2006). This is appropriate for NER because of the broad range of DNA damage it can repair. Fan *et al*., speculated that conserved domains of Rad25 recognise damage and because of the key role of Rad25 in TFIIH assembly this could act as a molecular switch for the alternative functions of TFIIH (Fan *et al.,* 2006).

The *H. volcanii* genome contains both eukaryotic and bacterial NER gene homologues (Hartman *et al*., in preparation) and there are four homologues of *RAD25-like* genes. *H. volcanii* is not unique in containing more than one *RAD25* homologue. Several crenarchaea contain two putative *rad25* genes (Kelman and White, 2005). *S. solfataricus* contains two copies of *rad25*/*xpb*, SsXpb-I is induced ~10-fold after UV exposure whereas SsXpb-II is constitutively expressed (Salerno *et al.,* 2003). These authors also tested the effect of the DNA damaging agent and transcription inhibitor actinomycin and found no change in expression for either gene. Like their eukaryotic homologues both SsXpb proteins are ssDNA-stimulated ATPases but neither has any helicase activity *in vitro* (Richards *et al.,* 2008). These authors suggest that eukaryotic and archaeal Rad25 proteins are not functional helicases *in vivo* but that their catalysis of ATP-dependent local DNA strand opening can look like helicase activity *in vitro*. This would support the model of Coin *et al.,* discussed earlier. Another example of multiple *rad25* genes is *Arabidopsis thaliana* which differs from other eukaryotes in that it contains two paralogous copies of *XPB* as the result of a recent tandem duplication (Costa *et al.,* 2001). Both copies are functional though there is some divergence in their roles. The *atrad25-1*$^{-/-}$ strain is viable which suggests that the atRad25-2 can complement atRad25-1 function (Costa *et al.,* 2001), (Morgante *et al.,* 2005).

To summarise, *RAD25* has a role in NER and transcription and defects result in UV sensitivity and a drastic reduction in transcription. Since the *H. volcanii rad25* homologues are associated with *oriC-2* and little is known about the role of Rad25 in archaea, these genes were characterised.

5.2 Bioinformatic analysis of *rad25* genes

The four *RAD25*-like helicases were named individually based on their homology with COG1061 (SSL2) which represents the superfamily II DNA/RNA helicases (Marchler-Bauer *et al.,* 2007). The four genes were named *rad25A-D* with *rad25A* showing the greatest homology and *rad25D* the least. Naming was based on BLAST scores using the September 2007 COG sequence and therefore differs from Table 5.2. *rad25D* is adjacent to *oriC-2* and *rad25B* and *rad25C* are also located near this replication origin.

Table 5.2: *rad25* homologues in *Haloferax volcanii*
Result of BLAST search querying the *H. volcanii* genome sequence (TIGR) with COG1061: SSL2 consensus sequence (see Appendix 7.3). The distance from *oriC-2* was calculated as the distance from the mid-point of the *rad25* gene to the mid-point of the intergenic region cloned in to pTA612 that functions as an origin of replication.

Gene	Replicon	ORF	BLAST score	Expect value	Gene size (kb)	Coordinates (bp)		Distance from *oriC-2* (kb)
						Start	Stop	
rad25A	pHV4	A00608	219	4×10^{-58}	1.40	449311	450711	n/a
rad25B	Chr	00206	213	5×10^{-56}	1.41	1607341	1608753	14.4
rad25C	Chr	00336	204	2×10^{-53}	1.38	1469248	1470627	123.7
rad25D	Chr	00213	188	1×10^{-48}	2.09	1590526	1592613	2.1

Figure 5.2 shows that there are a number of genes related to DNA processing (topoisomerase VI, gyrase and ligase) close to the *oriC-2* locus. There are also putative helicases near *rad25D* and *rad25B*. *rad25A* is on a different replicon to the other *rad25* homologues and is located close to several ISH51 elements. ISH51 is an insertion sequence-like element which is scattered across the *H. volcanii* genome and the genomes of other halophiles (Hofman *et al.,* 1986). pHV4 contains many ISH51 sequences as well as other sequences related to mobile genetic elements e.g. transposase. pHV4 has a drastically different codon bias to the main chromosome (S. Haldenby and S. Delmas unpublished observations) so it is reasonable to suggest that mobile elements significantly contributed to its organisation.

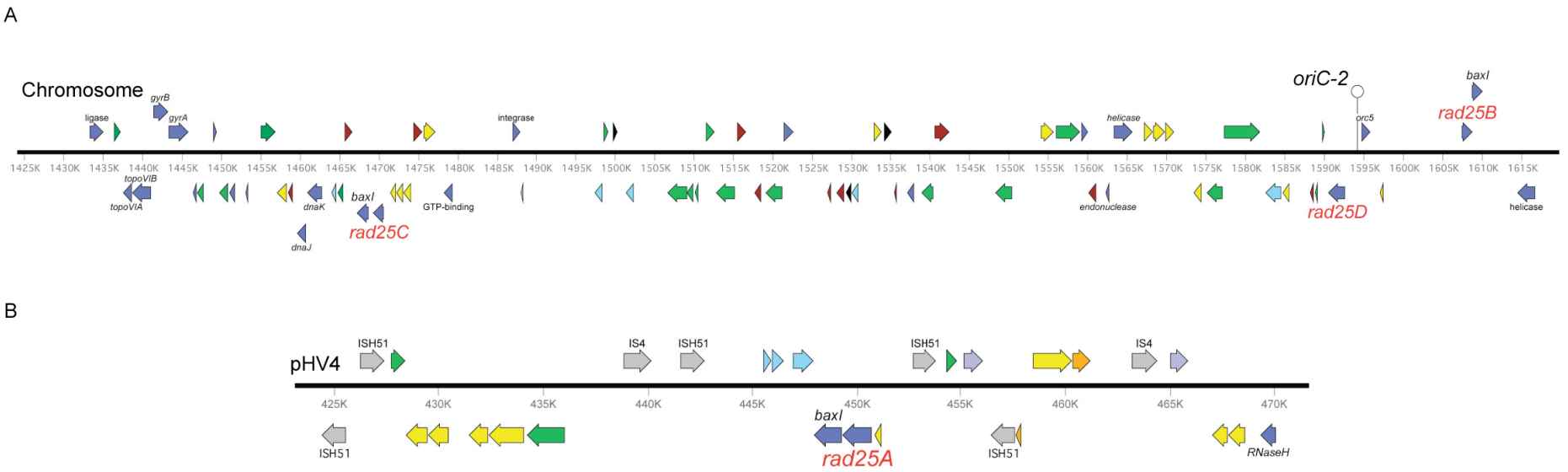


Figure 5.2: *rad25* gene loci in *Haloferax volcanii*

Unlabelled genes are coloured based on TIGR annotations. Yellow: transport, green: energy metabolism, black: transcription, grey: insertion sequence elements, blue: DNA processing, red: protein biosynthesis, purple: transcription, orange: stress response, light blue: protein fate. Hypothetical ORFs and those with unknown function have been removed for clarity.

(A) The *oriC-2* locus on the main chromosome with *rad25B*, *rad25C*, *rad25D* and *bax1* homologues shown (see text for details).

(B) The *rad25A* and *bax1* locus (see text for details), on pHV4.

In a search for nucleic acid cleaving enzymes, a family characterised by an endonuclease-like fold called DUF790 were identified in archaea and bacteria. Significantly every member was found to be adjacent to helicases of the SSL2 family (COG1061) (Kinch *et al.,* 2005). Richards *et al.*, proposed the name *bax1* (Binds archaeal XPB) for this gene because the purified *S. acidocaldarius* Bax1 interacts *in vitro* with *S. solfataricus* Xpb2 (Richards *et al.,* 2008). These authors found that *bax1* has a conserved association with *rad25* across the euryarchaea and crenarchaea. In crenarchaea with two *rad25* homologues one version was associated with *bax1* while the other did not have a partner. *H. volcanii* possesses three *bax1* homologues, each of which is associated with a *rad25* gene. *rad25D* is adjacent to *oriC-2* and has no associated *bax1* gene. *bax1* currently has no known function but the association of a putative nuclease and helicase has potential applications in NER.

Table 5.3: *bax1* homologues in *Haloferax volcanii*
Result of BLASTp search using *S. solfataricus bax1* on the *H. volcanii* genome

Replicon	ORF	Co-ordinates (bp)		E -value	Position relative to *rad25* genes
		Start	Stop		
Chr	ORF00338	1467212	1468714	2×10^{-38}	One gene from *rad25C*
pHV4	ORFA00609	447953	449311	2×10^{-33}	Adjacent to *rad25A*
Chr	ORF00205	1608740	1610092	2×10^{-29}	Adjacent to *rad25B*

The predicted domains of the Rad25 proteins were identified using the Pfam database (Finn *et al.,* 2008). Each contains a superfamily II helicase domain at the C-terminus and a type III restriction enzyme-like domain. Type III restriction enzymes are components of bacterial DNA restriction-modification mechanisms which protect the organism against invading foreign DNA. Rad25D is also predicted to contain an EspF protein proline-rich repeat. In *enteropathogenic E. coli*, EspF proteins are required for virulence and they induce host cell apoptosis . Since the E-value for the EspF protein repeat match is only just above the significance boundary, it is unlikely to be functionally relevant.

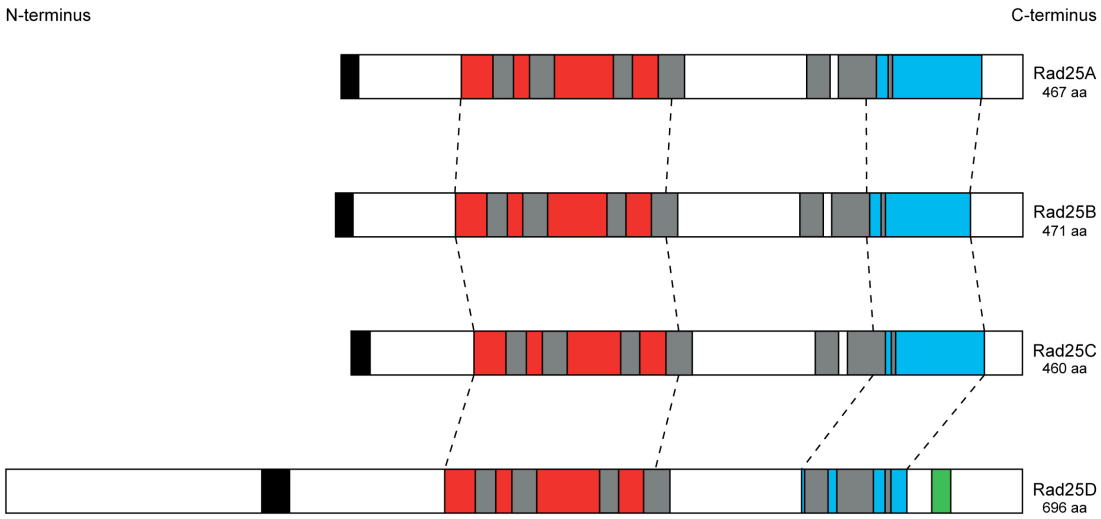


Figure 5.3: Rad25 domains in *Haloferax volcanii*
Significant predicted Pfam domain matches of the *H. volcanii* Rad25 proteins. Red denotes ResIII, a type III restriction enzyme-like domain and blue denotes the helicase domain. Rad25D also contains an EspF protein repeat match marked in green. The black box represents the DNA binding motif and the grey boxes show the helicase motifs (I, Ia and II-VI from left to right, for details see Figure 5.4).

The *H. volcanii* Rad25 proteins all contain a DNA binding domain and the seven helicase motifs, both characteristics of Rad25 homologues (Figure 5.3). The canonical Walker A and B boxes (domains I and II) are well conserved but the remaining five helicase domains are more degenerate. For example, domain VI is truncated when compared to the human or yeast version. This is similar to the *S. solfataricus* and *Halobacterium* sp. NRC-1 Rad25 homologues. Rad25A, Rad25B and Rad25C are similar but Rad25D is significantly larger due to extra N-terminal amino acids before the DNA binding domain. An alignment of the N-terminus with full length Rad25 from *S. cerevisiae* shows poor homology (Figure 5.5). In addition the DNA binding domain of Rad25D shows weak conservation compared to the other Rad25 homologues. BLAST analysis of the N-terminal domain produces significant matches with other putative archaeal DNA repair helicases but there are no significantly conserved domains within the N-terminus. In summary the N-terminus of Rad25D is conserved within the archaea but sequence analysis gives no clue to its function.

Bioinformatic analysis of the *H. volcanii* Rad25 proteins supports the annotation of these genes.

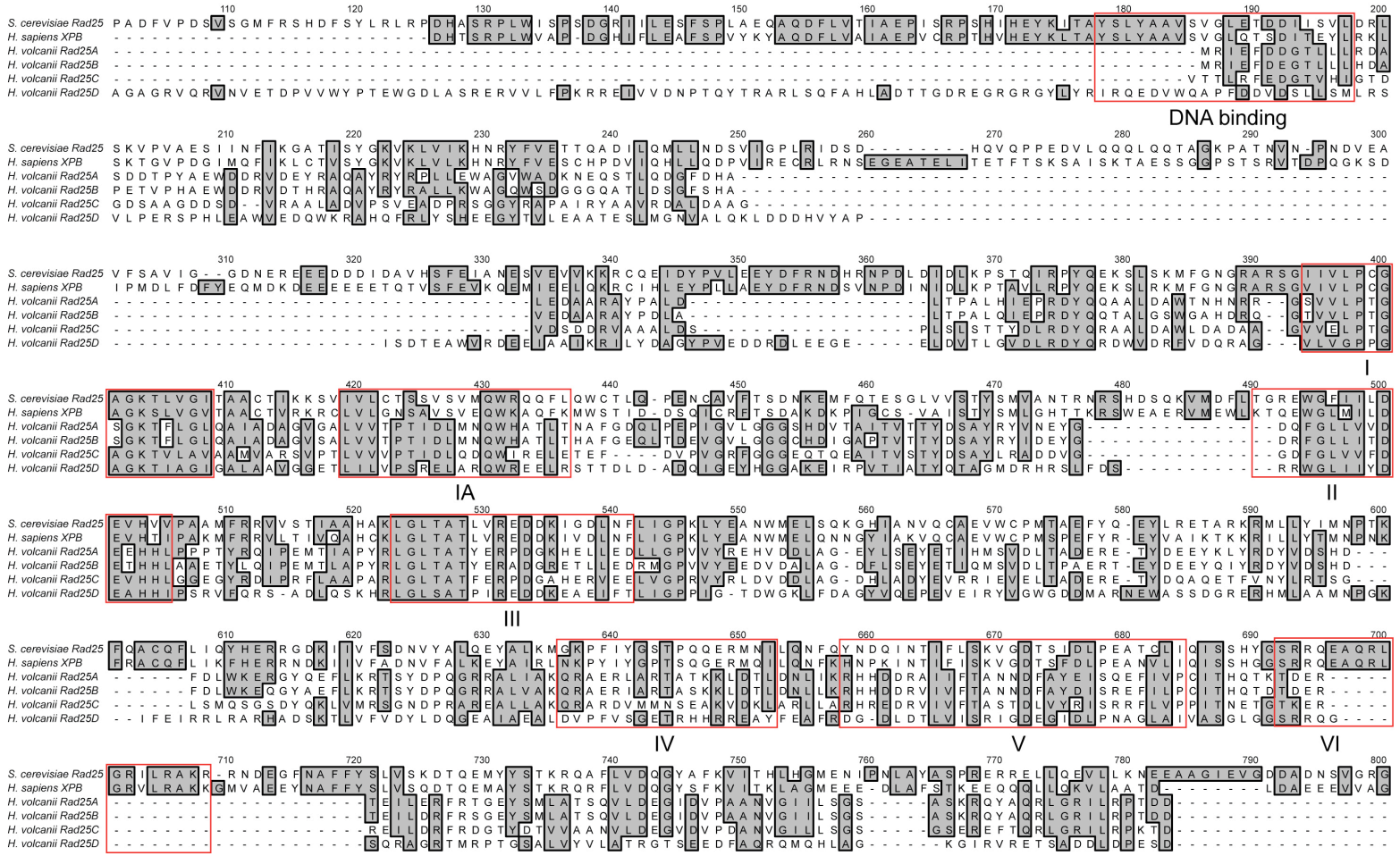


Figure 5.4: Helicase domains in *Haloferax volcanii* Rad25 proteins

Alignment of *H. sapiens* (accession number NM000122) and *S. cerevisiae* (accession number LO1414) Rad25 with the *H. volcanii* Rad25 proteins. The DNA binding domain and helicase domains are marked by boxes and labelled based on Morgante *et al*., 2004.

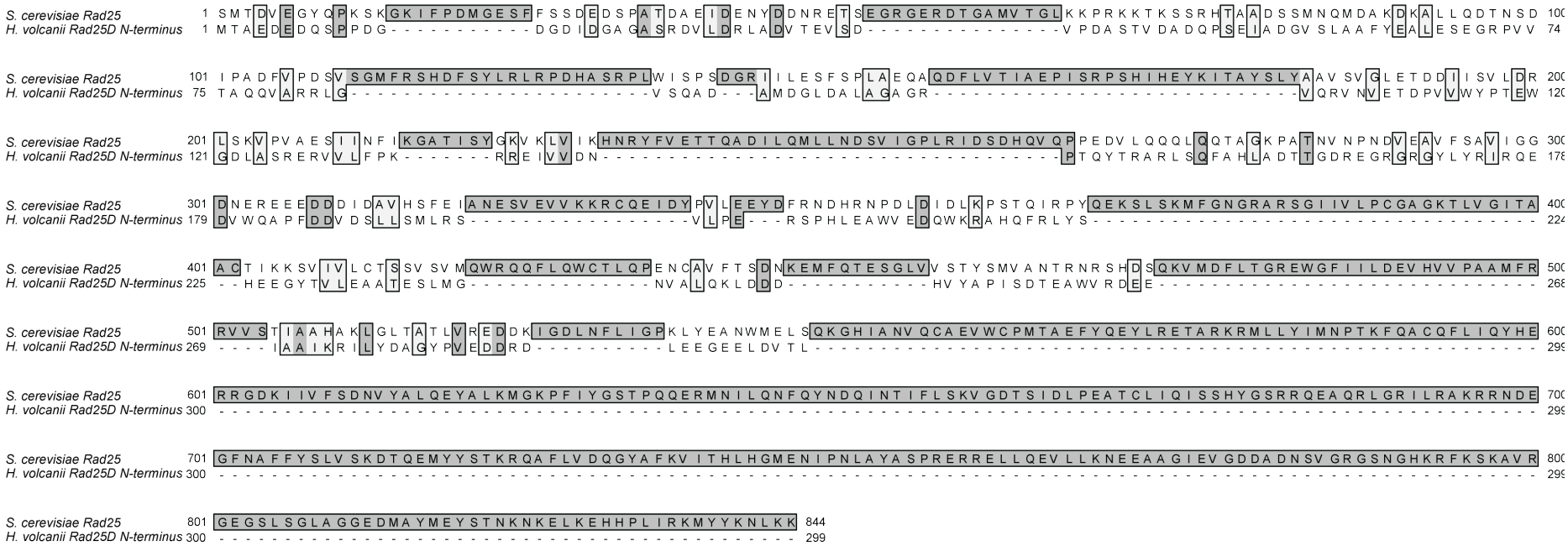


Figure 5.5: Rad25D N-terminus alignment

Alignment of *S. cerevisiae* (accession number Q00578) Rad25 with the *H. volcanii* Rad25 N-terminus. The N-terminus includes amino acids 1-299 and ends at the type III restriction enzyme-like domain boundary.

5.3 Construction of *rad25* deletion strains

rad25 deletion strains were constructed in order to characterise its role using genetics. A genomic clone of each *rad25* gene was isolated from H98 using the digests detailed in Table 5.4. The isolated fragments were ligated with pBluescript II SK+ (pTA35) cut with the same restriction enzymes.

Table 5.4: Genomic libraries for cloning *rad25* genes

Gene	Restriction enzyme(s)	Chromosomal fragment (kb)	Genomic clone	Deletion plasmid
rad25A	*Bam*HI/*Hin*dIII	5.2	pTA878	pTA882
rad25B	*Eco*RI	4.6	pTA725	pTA737
rad25C	*Kpn*I/*Xba*I	5	pTA731	pTA738
rad25D	*Not*I	9.3	pTA416	pTA453

Deletion constructs were made from the clones using either restriction enzyme digest and religation or by amplification of the flanking regions by PCR and ligation as shown in Figure 5.5. The deletion constructs were cloned in to pTA131 (*pyrE2*$^{+}$) using the inserted external restriction sites for *rad25A/D* shown in Figure 5.5. The deletion constructs for *rad25B/C* were subcloned in to pTA131 using *Eco*RI/*Not*I and *Kpn*I/*Xba*I respectively. The genomic clone and deletion construct plasmids for *rad25D* were made by Zhenhong Duan (University of Nottingham, Nottingham).

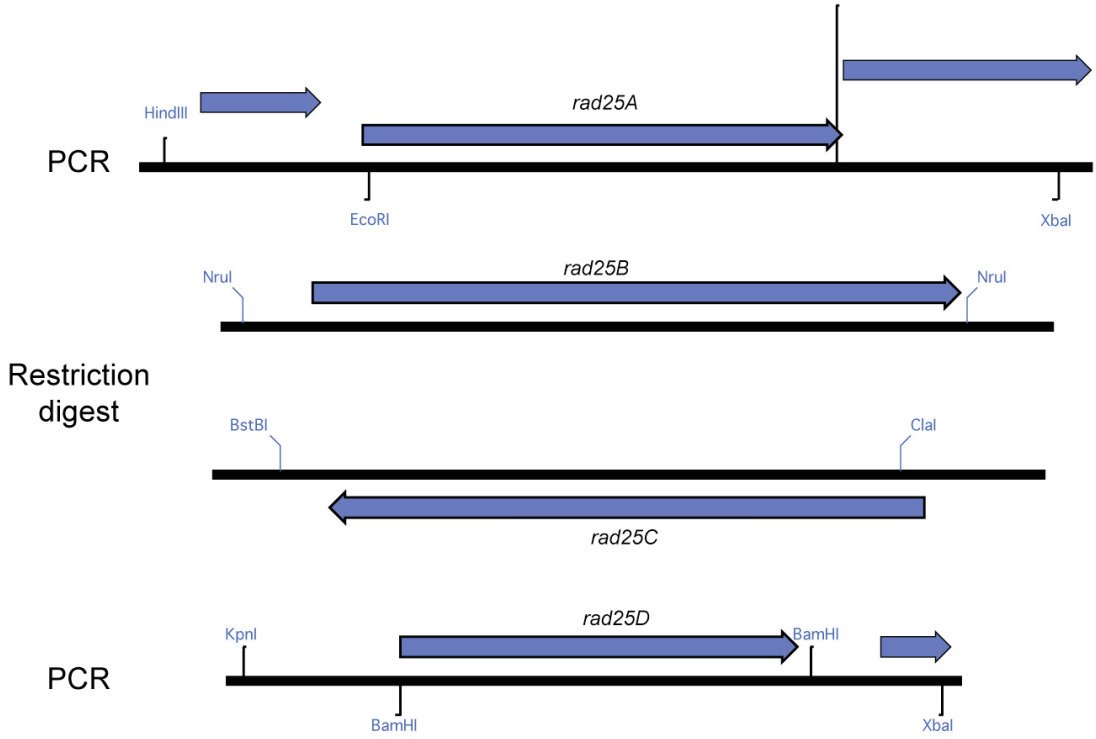


Figure 5.6: *rad25* gene deletion strategy

Deletion construct strategy for each *rad25* gene. *rad25B* and *rad25C* deletion constructs were made by cutting at the restriction sites shown and self-ligating. For *rad25A* and *rad25D* the primer positions and artificially inserted restriction sites are shown. The flanking regions were amplified by PCR, digested with *Eco*RI or *Bam*HI and ligated together to form the deletion construct.

The deletion constructs were used to transform the *ura*$^-$ *H. volcanii* strain H53 (Δ*pyrE2*, Δ*trpA*) to prototrophy for uracil. Integration was verified by restriction enzyme digest of genomic DNA and Southern blot analysis. Integrant strains were grown serially in Hv-YPC broth to relieve uracil selection and encourage pop-out of the integrated plasmid. Intramolecular recombinants that had lost the plasmid were counter selected by plating on Hv-Ca containing 5-FOA and tryptophan. The deletion mutants are *ura*$^-$ so the pop-in pop-out system could be re-used with the same marker. This method was used to construct the single, double, triple and quadruple *rad25* deletion strains.

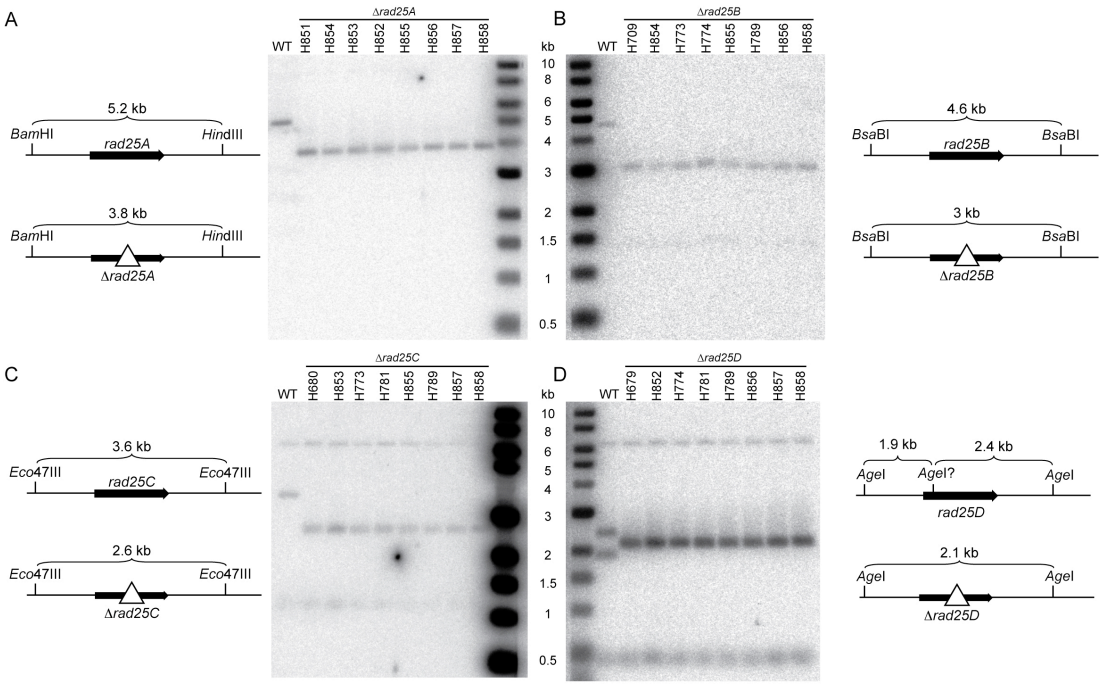


Figure 5.7: Southern blot analysis of *rad25* deletion strains

In all panels the wild-type (WT) is the parent strain H53 (Δ*pyrE2*, Δ*trpA*) and the diagnostic digest and predicted fragments are shown. The genotype of each strain is listed in Figure 5.7.

(A) Genomic DNA digested with *Bam*HI/*Hin*dIII and probed with the flanking regions of *rad25A* (pTA882 *Not*I/*Eco*RI).

(B) Genomic DNA digested with *Bsa*BI and probed with the flanking regions of *rad25B* (pTA737 *Not*I/*Eco*NI).

(C) Genomic DNA digested with *Eco*47III and probed with the flanking regions of *rad25C* (pTA738 *Xba*I/*Kpn*I).

(D) Genomic DNA digested with *Age*I and probed with the flanking regions of *rad25D* (pTA453 *Xba*I/*Kpn*I). The predicted wild-type fragment is 4.3 kb. There seems to be an *Age*I site within *rad25D* that is not present in the TIGR chromosome sequence. This extra site results in 1.9 and 2.4 kb bands for the wild-type locus and is deleted in the Δ*rad25D* strain.

It was possible to delete all the *rad25* genes individually and in combination, therefore their function is non-essential.

5.4 Growth phenotype of Δ*rad25* strains

All the *rad25* deletion strains were streaked on Hv-YPC to assess their growth. Colonies of the quadruple deletion strain are slightly smaller than wild-type but there were no drastic differences in growth on solid media for any of the deletion strains.

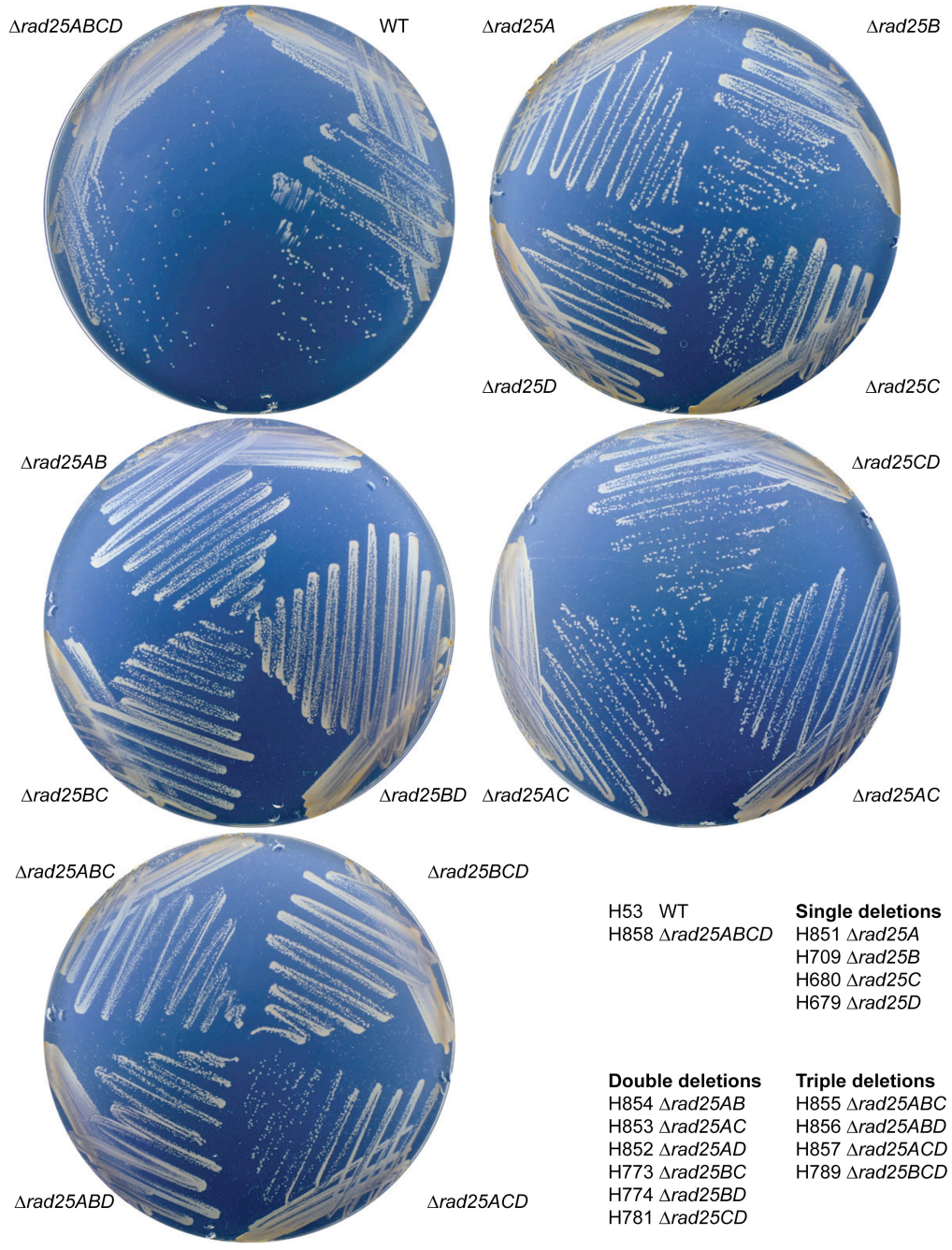


Figure 5.8: Δ*rad25* strain growth on solid media
Plate pictures of the single, double, triple and quadruple *rad25* deletion strains compared with wild-type on solid Hv-YPC media. Growth shown after four days of incubation at 45°C.

To quantify the growth difference between the Δ*rad25ABCD* and wild-type strains their generation time was measured.

Table 5.5: Generation time of Δ*rad25ABCD*
Generation time was estimated using the doubling time of viable cell numbers during exponential growth phase in Hv-YPC broth. SD = standard deviation. SE = standard error

Strain	Genotype	Generation time (hours)	n	SD	SE
H53	WT	2.8	3	0.047	0.03
H858	Δ*rad25ABCD*	3.8	2	0.21	0.15

An unpaired T-test on the generation times gives a P-value of 0.0036 so there is a statistically significant difference between growth of the wild-type and the Δ*rad25ABCD* strain. Therefore one or more of the *rad25* genes are important for normal growth.

5.4.1 Single Δ*rad25* deletion growth phenotype

To discern which gene(s) are responsible for the growth defect in Δ*rad25ABCD* the fitness of the individual *rad25* deletion strains was directly compared with the wild-type in paired competition assays. Prolonged simultaneous growth ensures that differences between wild-type and mutant strains are exaggerated due to direct competition. This enables subtle growth defects to be measured and is less subject to variables that can affect generation time measurements. To distinguish between wild-type and mutant cells during the assay, *β*-galactosidase was used as a marker. *bgaHa*$^+$ versions of the single and quadruple *rad25* deletion strains and the wild-type H53 strain were made by transformation with pTA102 (*pyrE2*, *bgaHa*$^+$). Correct integration was confirmed by growth on selective media and assaying with X-gal. Blue colonies were grown serially in Hv-YPC broth to relieve selection for uracil and encourage pop-out of pTA102. Putative pop-out colonies were sprayed with X-gal to identify *bgaHa*$^+$ strains. To avoid bias due to any cost of being *bgaHa*$^+$, each competition assay was performed with the *bgaHa*$^+$ marker on the wild-type strain and repeated with the *bgaHa*$^+$ marker on the Δ*rad25* strain.

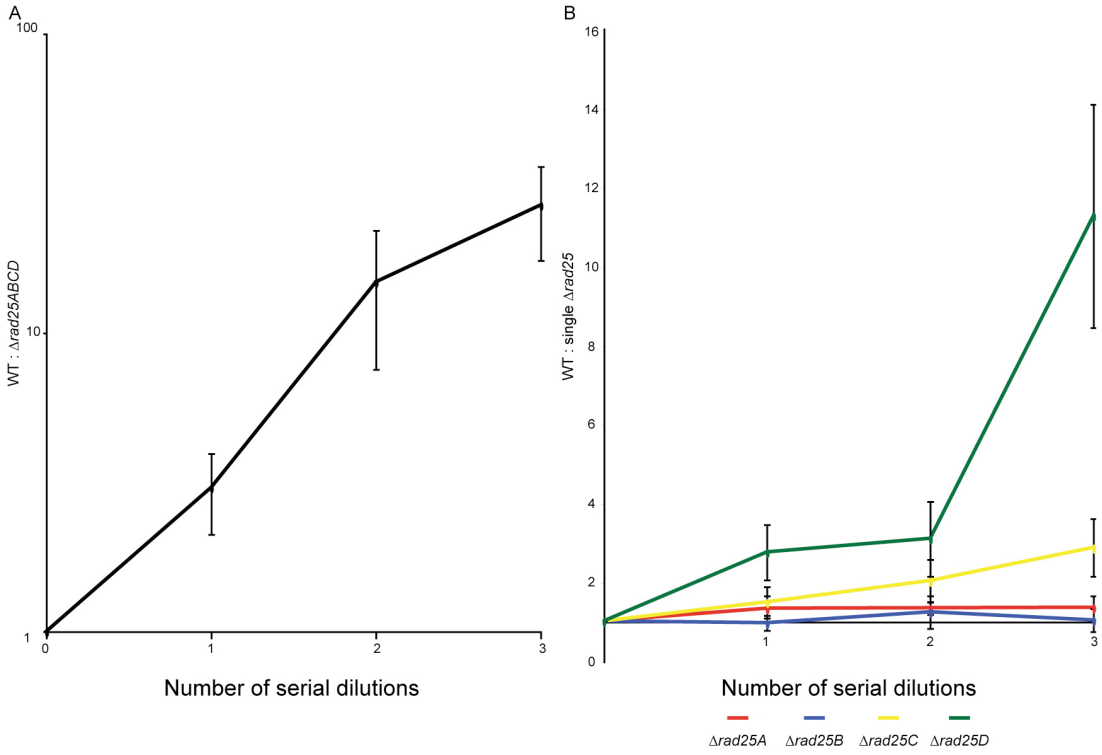


Figure 5.9: Δ*rad25* strain competition assays

(A) Average of five paired competition assays, two with wild-type H53 (Δ*pyrE2*, Δ*trpA*) and mutant H1009 (Δ*rad25ABCD*, *bgaHa*$^{+}$) and three with wild-type H1010 (*bgaHa*$^{+}$) and mutant H858 Δ*rad25ABCD*. Error bars represent one standard error.
(B) Average of four paired competition assays, two with wild-type H53 (Δ*pyrE2*, Δ*trpA*) and mutant H1008, H1007, H1006 or H1005 (Δ*rad25A/B/C/D*, *bgaHa*$^{+}$) and two with wild-type H1010 (*bgaHa*$^{+}$) and mutant H851, H709, H680 or H679 (Δ*rad25A/B/C/D*). Error bars represent one standard error. The x-axis is at y = 1 to indicate a 1:1 relationship

The paired competition assays between the wild-type and Δ*rad25ABCD* deletion strains showed that the wild-type rapidly out-competed the quadruple deletion strain. This recapitulates the generation time data and demonstrates that deleting all the *rad25* genes results in a growth defect. The paired competition assays using the single *rad25* deletion strains showed that Δ*rad25D* strains were significantly out-competed by the wild-type, Δ*rad25C* strains were slightly disadvantaged and Δ*rad25A/B* strains showed no significant difference from a 1:1 ratio. Deletion of *rad25D* alone is not responsible for the slow growth of Δ*rad25ABCD* because the ratios after three successive rounds of competition are different (27:1 compared to 11:1). Therefore the effect of the *rad25* genes on growth may be additive.

5.5 DNA damage sensitivity phenotype

To test the hypothesis that the *H. volcanii* Rad25 proteins are involved in DNA repair the Δ*rad25ABCD* strain was tested for DNA damage sensitivity. To test for activity of Rad25 in multiple damage repair pathways, mutagens that induce different types of DNA damage were used.

- UV radiation induces many types of DNA damage but 6-4 photoproducts and cyclobutane pyrimidine dimers which distort the DNA helix dominate (Friedburg, 2006). These lesions impede replication forks which can lead to double-strand breaks. A number of different repair mechanisms have evolved to repair UV-induced lesions, including photoreactivation, homologous recombination, BER and NER (Goosen and Moolenaar, 2008). To exclude direct reversal of the damage by photolyase activity (photoreactivation) UV-exposed *H. volcanii* were incubated in the dark (McCready, 1996).
- Methyl methanesulfonate (MMS) is an alkylating agent that reacts with a nucleophilic centre in DNA to add methyl groups to guanine and adenine bases (Wyatt and Pittman, 2006). Methyl groups are removed by BER and alkyltransferase action (Lindahl and Wood, 1999).
- Mitomycin C (MMC) is an interstrand cross-linker (Iyer and Szybalski, 1964). Covalently linked complementary strands can block transcription and replication and are particularly deleterious because both strands need repairing. Repair utilises many different pathways including NER, homologous recombination and translesion synthesis (NER and DNA polymerase II) (reviewed in Noll *et al.,* 2006).
- The products of hydrogen peroxide break down in the iron-mediated Fenton reactions cause oxidative base damage and strand breaks (Henle and Linn, 1997). Multiple repair processes are responsible for repair of this damage with BER being the most common pathway (Slupphaug *et al.,* 2003).

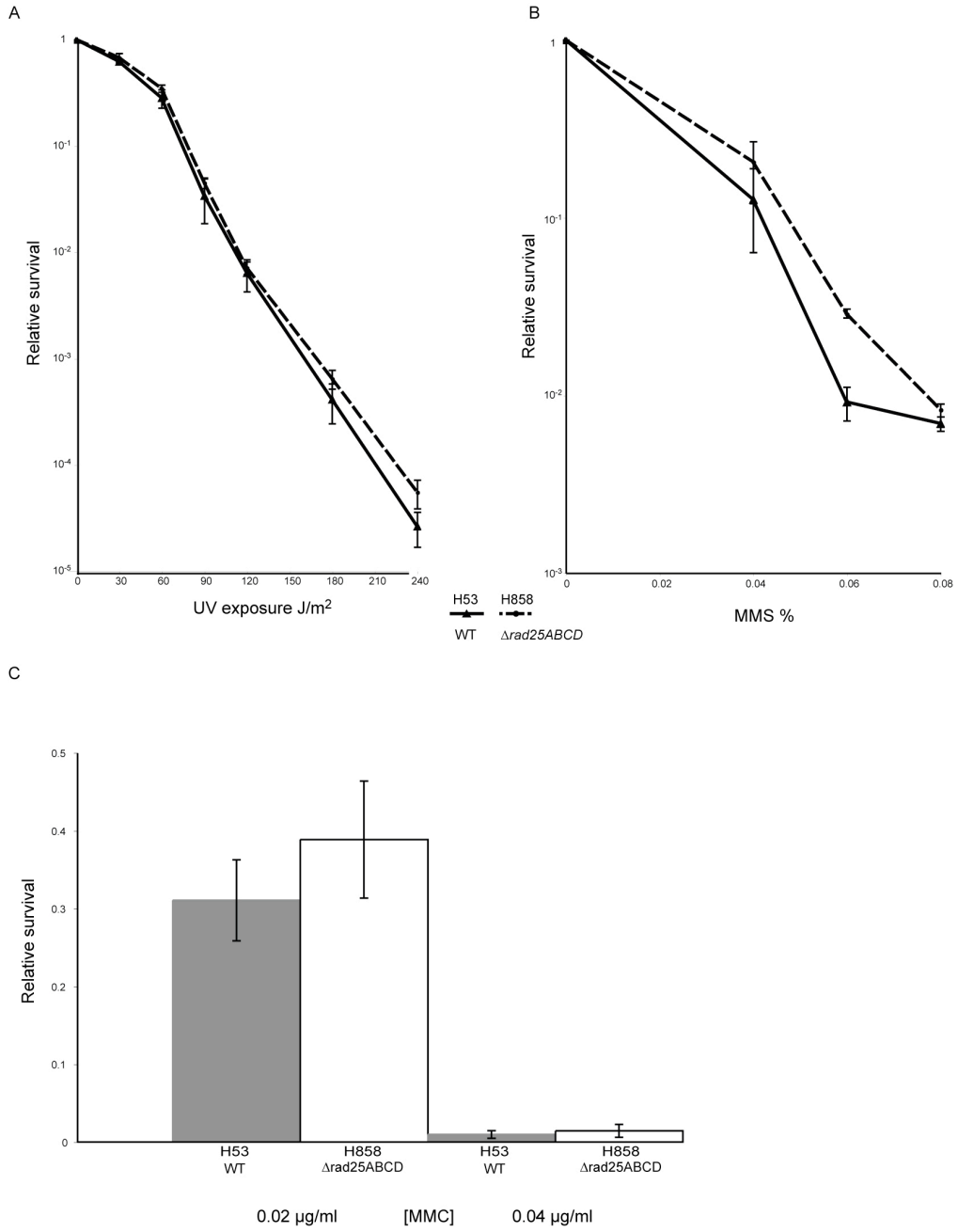


Figure 5.10: Δ*rad25ABCD* sensitivity to UV, MMS and MMC exposure

(A) Survival of mutant H858 (Δ*rad25ABCD*) and wild-type H53 (Δ*pyrE2*, Δ*trpA*) following UV exposure. Data points are the mean of three experiments, error bars represent one standard error.

(B) Survival of mutant H858 (Δ*rad25ABCD*) and wild-type H53 (Δ*pyrE2*, Δ*trpA*) following 1 hour of MMS exposure. Data points are the mean of two experiments, error bars represent one standard error.

(C) Survival of mutant H858 (Δ*rad25ABCD*) and wild-type H53 (Δ*pyrE2*, Δ*trpA*) following growth on solid media containing MMC. Data points are the mean of three experiments, error bars represent one standard error.

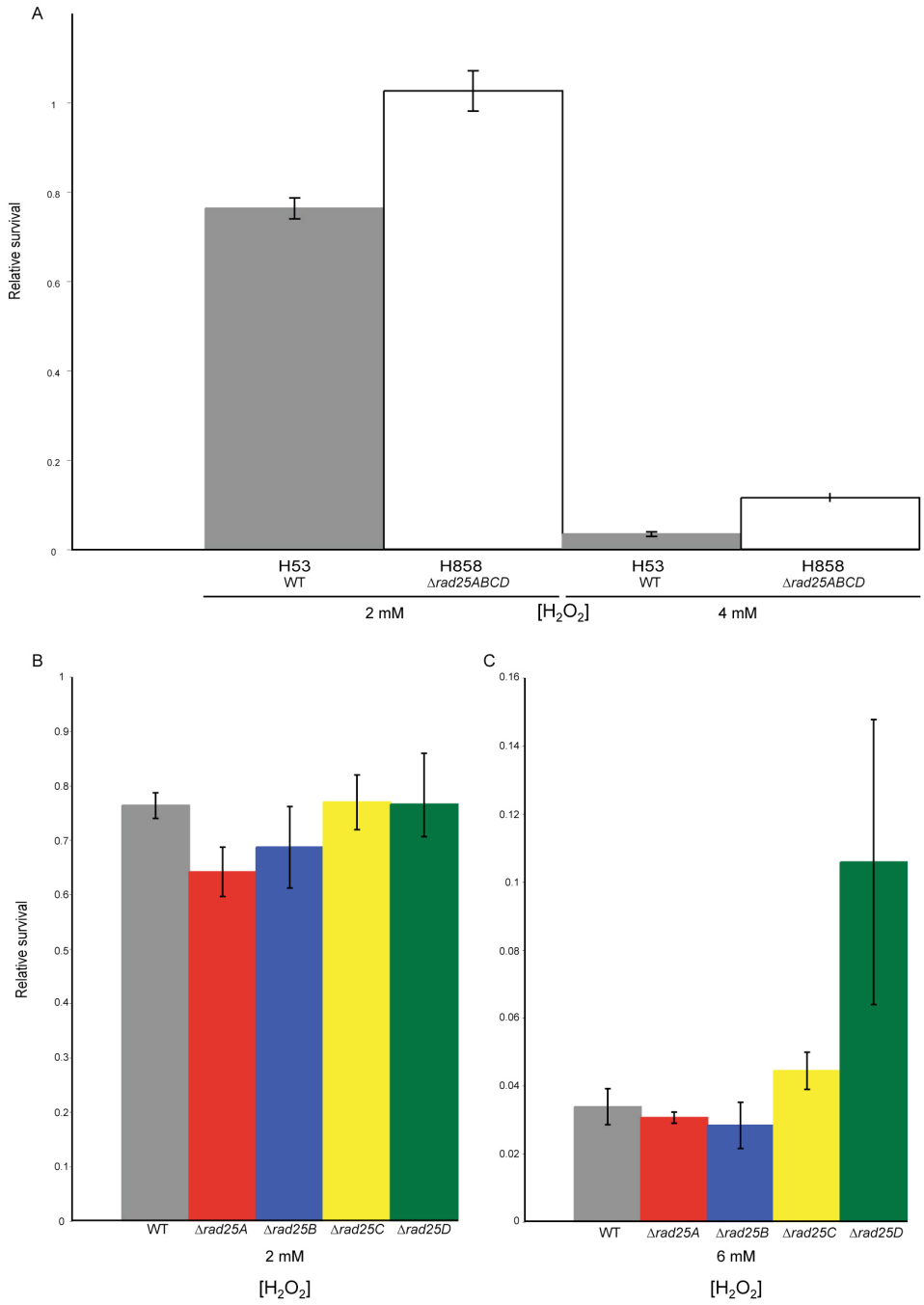


Figure 5.11: Δ*rad25ABCD* and single *rad25* deletion strains sensitivity to hydrogen peroxide exposure

(A) Survival of mutant H858 (Δ*rad25ABCD*) and wild-type H53 (Δ*pyrE2*, Δ*trpA*) following 1 hour of hydrogen peroxide exposure. Data points are the mean of three experiments, error bars represent one standard error.

(B) and (C) Survival of H851 (Δ*rad25A*), H709 (Δ*rad25B*), H680 (Δ*rad25C*), H679 (Δ*rad25D*) and wild-type H53 (Δ*pyrE2*, Δ*trpA*), following 1 hour of hydrogen peroxide exposure. Data points are the mean of two experiments, error bars represent one standard error.

Figure 5.9 shows that there is no increased sensitivity to UV radiation, MMS or MMC exposure when all the *rad25* genes are deleted in *H. volcanii*. Figure 5.10 shows that deletion of all the *rad25* genes gives a survival advantage after hydrogen peroxide-induced DNA damage. This advantage is significant at 2 mM hydrogen peroxide in the Δ*rad25ABCD* strain (Figure 5.10, panel A), but when the single deletions were examined there were no gross differences in survival. This could suggest that the four Rad25 proteins confer this hydrogen peroxide resistance in an additive manner. In contrast, with 6 mM hydrogen peroxide exposure Δ*rad25D* confers a significant advantage while the other deletion strains show the same survival as the wild-type. Given the competition assays in Figure 5.8, panel B, it is likely that the advantage of being Δ*rad25D* is indirect and the result of its longer generation time. This would give cells longer to repair DNA damage. The three-fold advantage of Δ*rad25D*, slight advantage of Δ*rad25C* and no difference for Δ*rad25A*/*B* shown in Figure 5.10, panel C; recapitulate the relative differences these strains showed in the competition assays (Figure 5.8, panel B). From these results it appears that the *rad25* genes do not function in DNA repair pathways.

5.6 Rad25 and *oriC-2* function

Given that *rad25D* and *rad25B* are located near *oriC-2*, another hypothesis for Rad25 function is that it aids duplex melting through helicase activity as a prelude to *oriC-2* firing. To test this hypothesis, *radA* was deleted in the Δ*rad25ABCD* background so that the ARS assay could be used to test *oriC-2* based plasmids. If *oriC-2* activity is diminished there will be fewer transformants because the episomes are not replicated efficiently. *radA* was also deleted in all the single, double and triple *rad25* deletion backgrounds to enable assays for the role of each individual gene if a phenotype was forthcoming. Despite Chapter 4 showing that *radA*-independent recombination of *oriC-2* sequences is possible, the ARS assay is still a valid test of this hypothesis because ample copies of the *oriC-2* based plasmid remain episomal when *radA*-independent recombination occurs (Figure 4.19, panel B).

To delete *radA* in the *rad25* deletion backgrounds a complementation approach was taken (Figure 5.11, panel A). *H. volcanii* is polyploid so each chromosome copy has to undergo intermolecular recombination in order to delete every copy of a gene. *radA* is the recombinase required for intramolecular recombination and functions as a polymer. Therefore when deleting *radA* it becomes progressively more difficult for chromosomes to recombine and pop-out the integrated plasmid due to the low gene dosage of *radA*. Complementation with an episomal source of *radA* ensures there is ample RadA to accomplish the final stages of plasmid pop-out (Delmas *et al*., submitted). This method improves the chance of obtaining a *radA* deletion. Δ*rad25* strains were transformed with a marked Δ*radA* construct on pTA324 (Δ*radA*::*trpA*, *pyrE2*). Transformants were selected on Hv-Ca and checked by restriction digest and Southern blotting. Integrants were transformed with pTA636 (*radA*$^{+}$, *pyrE2*, Mev^{R}) as the *in trans* supply of RadA. Transformants were selected on Hv-Ca+mev. Cultures were propagated twice to stationary phase while under selection for mevinolin resistance and tryptophan prototrophy but with selection relieved for uracil (Hv-Ca broth +ura). Then selection was relieved for mevinolin resistance during serial growth to cure the plasmid. At all times selection was maintained for the Δ*radA*::*trpA* genotype by using Hv-Ca media. After three successive serial dilutions, cells that had lost the integrated and cured plasmid were counter-selected

using 5-FOA. These plates were incubated at 45°C for at least 10 days because Δ*radA* strains show a significant growth defect. Small colonies were patched and checked for Δ*radA* deletion by probing colony lift membranes with part of the deleted *radA* gene (pTA29 *Nco*I/*Not*I). Colonies that did not hybridise with the probe were confirmed as Δ*radA* by restriction digestion of genomic DNA and Southern blotting (Figure 5.11, panel C).

It was possible to delete *radA* in all the *rad25* deletion strains. To complement Figure 5.11, Table 5.6 lists the strain backgrounds in which *radA* was deleted.

Table 5.6: Strain backgrounds of Δ*radA* deletions

Strain number	*rad25* genotype	Strain number	*rad25* genotype
H990	Δ*rad25A*	H830	Δ*rad25B*
H819	Δ*rad25C*	H831	Δ*rad25D*
H909	Δ*rad25A* Δ*rad25B*	H933	Δ*rad25A* Δ*rad25C*
H908	Δ*rad25A* Δ*rad25D*	H821	Δ*rad25B* Δ*rad25C*
H822	Δ*rad25B* Δ*rad25D*	H833	Δ*rad25C* Δ*rad25D*
H991	Δ*rad25A* Δ*rad25B* Δ*rad25C*	H992	Δ*rad25A* Δ*rad25B* Δ*rad25D*
H934	Δ*rad25A* Δ*rad25C* Δ*rad25D*	H834	Δ*rad25B* Δ*rad25D* Δ*rad25C*
H910	Δ*rad25A* Δ*rad25B* Δ*rad25C* Δ*rad25D*		

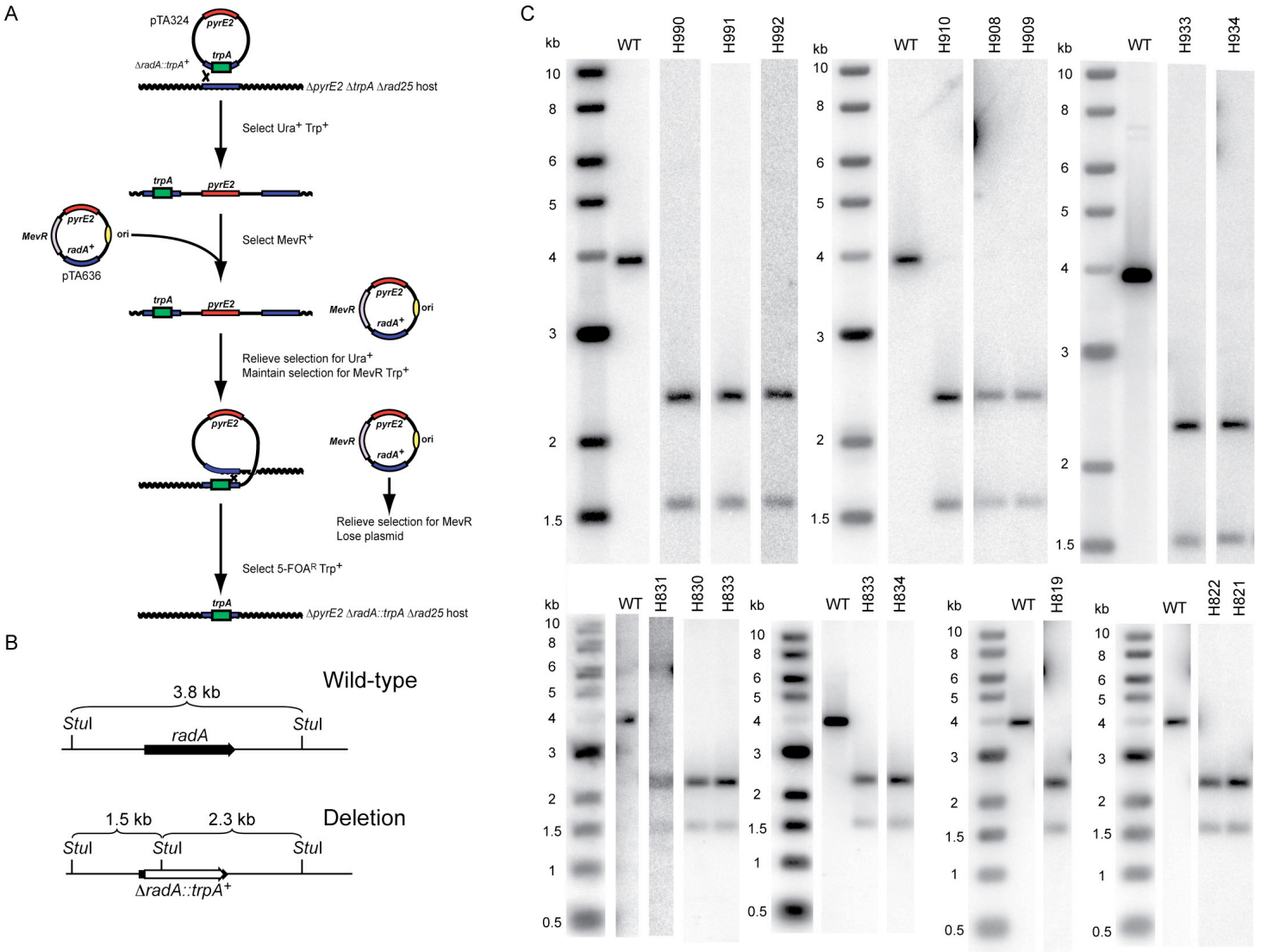


Figure 5.12: Deletion of *radA* in various Δ*rad25* backgrounds

(A) *radA* deletion strategy in Δ*rad25* strain backgrounds using *radA in trans* on a Mev^R complementing plasmid to increase pop-out likelihood.

(B) Diagnostic digest and predicted fragments for Δ*radA*::*trpA*.

(C) Southern blots showing *radA* deletion. Genomic DNA was digested with *Stu*I and hybridised with pTA29 *Kpn*I/*Eco*RI which contains the *radA* gene and its flanking regions.

The Δ*radA*, Δ*rad25ABCD* (H910) strain was used to test *oriC-2* based plasmids in an ARS assay. In addition to testing the 2.4 kb intergenic *oriC-2* region (pTA612), a larger clone of the *oriC-2* locus was used (pTA419) which includes *oriC-2* and the adjacent *rad25D* gene. The purpose of this was to test whether an episomal supply of *rad25D* produced a different result from the lack of any *rad25* gene products in the Δ*rad25ABCD* background.

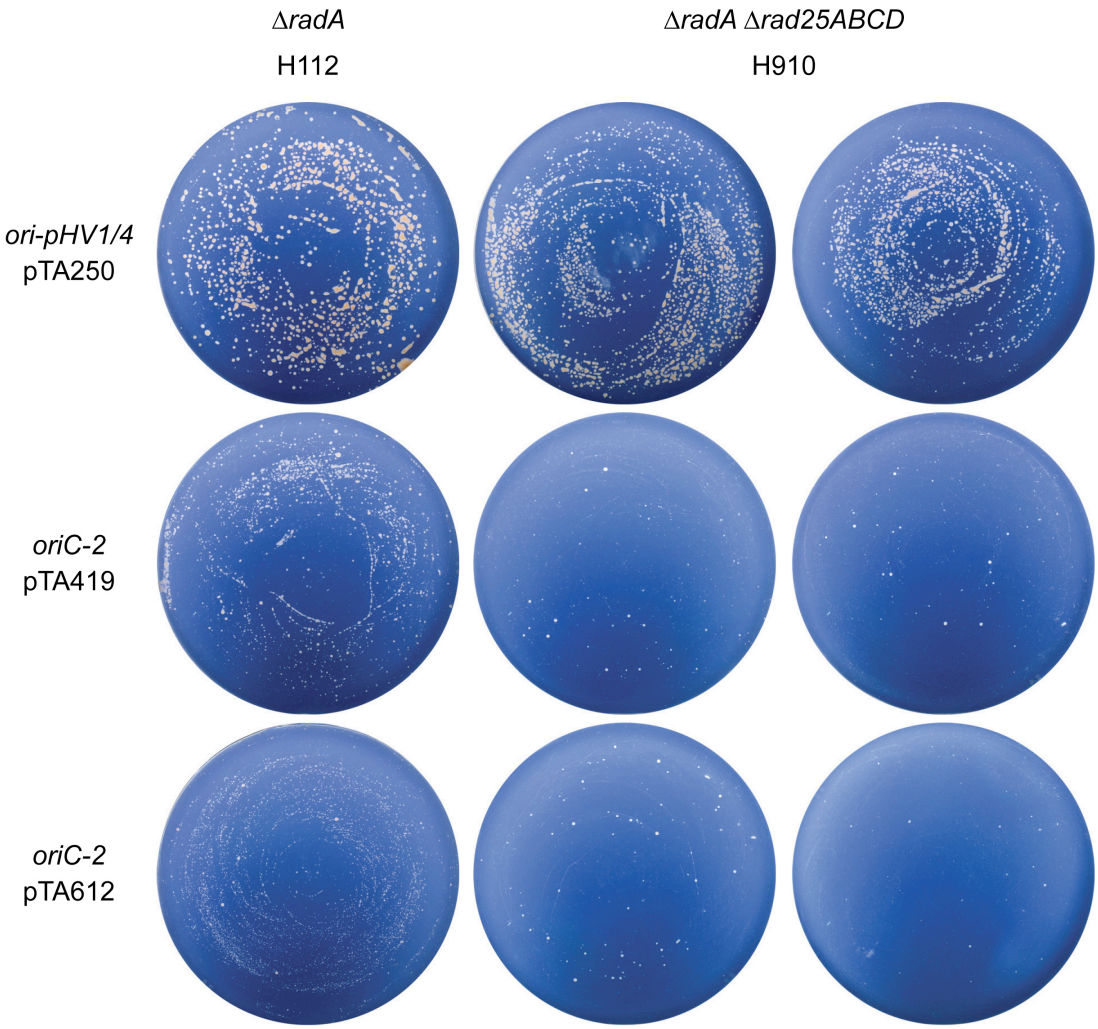


Figure 5.13: Δ*rad25ABCD* transformed with *oriC-2* based plasmids
H112 (Δ*radA*, Δ*pyrE2*) and H910 (Δ*radA::trpA*, Δ*pyrE2,* Δ*rad25ABCD*) transformed with 1 µg of pTA250 (*pyrE2*, *ori-pHV1/4*), pTA419 (*oriC-2*) or pTA612 (*oriC-2*). pTA419 is a 5.5 kb *oriC-2* clone that also contains *rad25D* while pTA612 contains intergenic *oriC-2*. The middle plate column represents experiment 1 and the right column represents experiment 2. Growth shown after 12 days at 45°C.

Table 5.7: Transformation efficiencies in a Δ*rad25ABCD* background
Efficiency is calculated by dividing the number of colonies on selective plates by the number of viable cells on non-selective plates for 1 µg of transforming DNA.

Plasmid	Strain		
	Δ*radA*	**Δ*rad25ABCD*, Δ*radA***	
		Experiment 1	**Experiment 2**
ori-pHV1/4	3.4%	1.7%	3.3%
oriC-2 + *rad25D*	2.2%	0.07%	0.06%
oriC-2 (intergenic)	1.8%	0.03%	0.05%

ori-pHV1/4 was used as a positive control to check that Δ*rad25ABCD* can be transformed as well as a *rad25*$^{+}$ strain. The results showed variability in the transformation efficiency of *ori-pHV1/4*; while experiment 2 showed normal efficiency, the transformation efficiency of experiment 1 was reduced by 50%. Taking in to account a potential reduction of 50%, qualitative assessment of Figure 5.12 shows that the *oriC-2* transformation efficiencies of the Δ*rad25ABCD* strain are much lower than expected when compared to the Δ*radA* strain transformation efficiency. This supports the hypothesis that Rad25 might help *oriC-2* activity. If the Rad25 proteins have a function in *oriC-2* firing then pTA419 (*oriC-2*, *rad25D*) might be expected to transform Δ*rad25ABCD* more efficiently than pTA612 (*oriC-2*) due to the episomal copy of *rad25D*. These results do not demonstrate this. Given the variability of results in Table 5.7, this can be interpreted only as preliminary evidence of a role for Rad25 in *oriC-2* function.

5.7 Discussion

The quadruple Δ*rad25* strain is viable so the *H. volcanii rad25* genes are non-essential. The minor growth defect of this strain indicates that one or more of the Rad25 proteins functions in normal growth. Preliminary data suggesting a possible role for Rad25 in *oriC-2* activity is worth pursuing. A possible mechanism is Rad25 helicase activity helping to unwind the DUE before the loading of replication factors. If this were the case then Δ*rad25ABCD* strains would lack duplex remodelling at the episomal and chromosomal *oriC-2* locus. Lack of episomal plasmid would explain the lower transformation efficiency of the *oriC-2*-based plasmids. Slower replication of the main chromosome due to impaired *oriC-2* firing could explain the longer generation time of the Δ*rad25ABCD* strain. It is also possible that Rad25 acts at other replication origins. This is an attractive model but it is not supported by Rad25 work in other organisms. *S. solfataricus* Rad25 does not show helicase activity *in vitro* and the prevailing opinion about eukaryotic NER is that Rad25-DNA binding and ATPase activity is crucial rather than any Rad25-mediated unwinding.

Given the variability of *ori-pHV1/4* transformation efficiency, a priority should be to determine the ability of the Δ*rad25ABCD*, Δ*radA* strain to take up DNA. Unfortunately potential experiments are limited because integrative plasmids can not be used due to the recombinase deletion. Plasmids based on various *H. volcanii* replication origins isolated in this work could be used but it is possible that Rad25 could also influence their activity. An alternative would be to use *ori-pHV2* based plasmids but unfortunately these do not function in Δ*radA* strains (Woods and Dyall-Smith, 1997). Addition of episomal *rad25D* (using pTA419) did not significantly improve the transformation efficiency of Δ*rad25ABCD*. This could be because insufficient Rad25D is produced from the episomal plasmid due to its low copy number. Alternatively Rad25 function may be cooperative and require all four proteins. A key test would be to add back the *rad25* genes on a high copy number plasmid and see if there is any change in Δ*rad25ABCD* transformation efficiency. This could be done in conjunction with conducting this ARS assay in the various single, double and triple *rad25*, *radA* deletion strains to look for gene-specific effects. Another consideration is the degree to which *radA*-independent recombination affected these results. It would be interesting to test whether this pathway is affected by the lack of Rad25 and/or reduced origin firing but to

accomplish this a quantitative assay to follow *radA*-independent recombination would need to be developed.

Unlike eukaryotic *RAD25*, it appears that the *H. volcanii rad25* genes are not involved in DNA repair because the Δ*rad25ABCD* strain shows no increase in DNA damage sensitivity. Specifically, the lack of a UV sensitive phenotype indicates that Rad25 has no role in NER and is therefore not the functional homologue of XPB/ERCC-1. *H. volcanii* also contains homologues of bacterial NER genes and strains deleted for *uvrA* are extremely sensitive to UV radiation exposure (Duan, 2009). Therefore it is likely that the majority of UV-induced DNA damage in *H. volcanii* is repaired via the UvrABC pathway. This is similar to the UvrABC system requirement for UV-induced damage repair for the closely related *Halobacterium* sp. NRC-1 which also contains homologues of eukaryotic NER proteins (Crowley *et al.,* 2006). It is possible that the eukaryotic NER homologues have a role in a back up DNA damage repair pathway which functions in the absence of *UvrABC*. To test this hypothesis *uvrA* would need to be deleted in the Δ*rad25ABCD* background. It is possible that any *rad25*-mediated repair is rare and/or conditional and therefore difficult to identify through this type of genetic analysis. There are also other DNA damaging agents that could be tested (e.g. cisplatin, phleomycin) which might produce DNA damage repaired through any *rad25*-dependent pathways. This work does not completely exclude the possibility that *rad25* has a role in DNA repair, but it is unlikely.

Since there is no direct evidence for a eukaryotic-like NER pathway in archaea it is possible that the eukaryotic NER homologues carry out alternative functions rather than an archaeal version of NER. The only archaeal components of TFIIH that have been identified are Rad3 and Rad25 so it is unlikely that archaea have a similar complex. It is possible that archaea have an unidentified analogous complex, however this is unlikely because *in vitro* reconstitution of transcription has shown that TFIIH function is not required in the archaea (Hethke *et al.,* 1996; Bell *et al.,* 1999). Expression analysis of the *rad25* and *bax1* genes might give some clues to their function. Of the four *H. volcanii rad25* genes *rad25D* is atypical in that it has an extended N-terminus, a poorly conserved DNA binding domain, no *bax1* partner and the greatest phenotypic effect on generation time when deleted. Therefore *rad25D* might act in a different pathway and should be considered separately.

Future perspectives

Archaea can serve as a simplified model of eukaryotic DNA replication. The identification and characterisation of *H. volcanii* replication origins has established the framework for further study of DNA replication in this species. This work has described the first example of multiple replication origins on a single replicon in the euryarchaeota. In the immediate future the priority should be to fully describe *oriC-3*, it will also be important to map any remaining unrecognised origins. Genetic studies can then begin to focus on the regulation of multireplicon replication. Key questions include determining how multiple replication origins on the main chromosome are regulated and how replication proceeds in the context of the archaeal cell cycle. Development of a method to synchronise *H. volcanii* and a quantitative *in vivo* assay of origin activity would be of great benefit in addressing these issues.

A unifying theme of *H. volcanii* biology is the impact of its dynamic genome. This has been evident in this work by the observance of replication origin translocations and the mobility of pHV4 sequences. The importance of these rearrangements in the evolution of its multireplicon genome structure has yet to be determined. The existence of an origin-specific recombination pathway and the presence of multiple replication origins suggest that origin duplication and translocation may have been a driving force in the evolution of *H. volcanii* genome organisation.

Chapter 6: Bibliography

Aladjem MI, Groudine M, Brody LL, Dieken ES, Fournier RE, Wahl GM, Epner EM (1995) Participation of the human beta-globin locus control region in initiation of DNA replication. *Science* **270:** 815-819

Aladjem MI, Rodewald LW, Kolman JL, Wahl GM (1998) Genetic dissection of a mammalian replicator in the human beta-globin locus. *Science* **281:** 1005-1009

Alberts B (2003) DNA replication and recombination. *Nature* **421:** 431-435

Allers T, Mevarech M (2005) Archaeal genetics - the third way. *Nat Rev Genet* **6:** 58-73

Allers T, Ngo H, Mevarech M, Lloyd RG (2004) Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl Environ Microbiol* **70:** 943-953

Aravind L, Koonin EV (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res* **27:** 4658-4670

Asai T, Bates DB, Boye E, Kogoma T (1998) Are minichromosomes valid model systems for DNA replication control? Lessons learned from *Escherichia coli*. *Mol Microbiol* **29:** 671-675

Austin RJ, Orr-Weaver TL, Bell SP (1999) Drosophila ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes Dev* **13:** 2639-2649

Bach T, Morigen, Skarstad K (2008) The initiator protein DnaA contributes to keeping new origins inactivated by promoting the presence of hemimethylated DNA. *J Mol Biol* **384:** 1076-1085

Baliga NS, Bonneau R, Facciotti MT, Pan M, Glusman G, Deutsch EW, Shannon P, Chiu Y, Weng RS, Gan RR, Hung P, Date SV, Marcotte E, Hood L, Ng WV (2004) Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res* **14:** 2221-2234

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289:** 905-920

Barry ER, Bell SD (2006) DNA replication in the archaea. *Microbiol Mol Biol Rev* **70:** 876-887

Bell L, Byers B (1979) Occurrence of crossed strand-exchange forms in yeast DNA during meiosis. *Proc Natl Acad Sci U S A* **76:** 3445-3449

Bell L, Byers B (1983) Separation of branched from linear DNA by two-dimensional gel electrophoresis. *Anal Biochem* **130:** 527-535

Bell SD, Kosa PL, Sigler PB, Jackson SP. (1999) Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci U S A* **96**:13662-7.

Bell SP (2002) The origin recognition complex: from simple origins to complex functions. *Genes Dev* **16:** 659-672

Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* **71:** 333-374

Bell SP, Mitchell J, Leber J, Kobayashi R, Stillman B (1995) The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* **83:** 563-568

Bell SP, Stillman B (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357:** 128-134

Berdichevsky A, Izhar L, Livneh Z (2002) Error-free recombinational repair predominates over mutagenic translesion replication in *E. coli*. *Mol Cell* **10:** 917-924

Berquist BR, DasSarma P, DasSarma S (2007) Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. *BMC Genet* **8:** 31

Berquist BR, DasSarma S (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. *J Bacteriol* **185:** 5959-5966

Bielinsky AK, Gerbi SA (1998) Discrete start sites for DNA synthesis in the yeast ARS1 origin. *Science* **279:** 95-98

Bitan-Banin G, Ortenberg R, Mevarech M (2003) Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *J Bacteriol* **185:** 772-778

Bohlke K, Pisani FM, Rossi M, Antranikian G (2002) Archaeal DNA replication: spotlight on a rapidly moving field. *Extremophiles* **6:** 1-14

Branzei D, Foiani M (2007) Interplay of replication checkpoints and repair proteins at stalled replication forks. *DNA Repair (Amst)* **6:** 994-1003

Braun RE, O'Day K, Wright A (1985) Autoregulation of the DNA replication gene dnaA in E. coli K-12. *Cell* **40:** 159-169

Breslauer KJ, Frank R, Blocker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci U S A* **83:** 3746-3750

Breuert S, Allers T, Spohn G, Soppa J (2006) Regulated polyploidy in halophilic archaea. *PLoS ONE* **1:** e92

Brewer BJ, Fangman WL (1987) The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* **51:** 463-471

Brewer BJ, Fangman WL (1988) A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell* **55:** 637-643

Brewer BJ, Fangman WL (1993) Initiation at closely spaced replication origins in a yeast chromosome. *Science* **262:** 1728-1731

Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6:** 245-252

Brukner I, Sanchez R, Suck D, Pongor S (1995) Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *Embo J* **14:** 1812-1818

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NSM, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273:** 1058-1073

Burhans WC, Vassilev LT, Caddle MS, Heintz NH, DePamphilis ML (1990) Identification of an origin of bidirectional DNA replication in mammalian chromosomes. *Cell* **62:** 955-965

Cabrera JA, Bolds J, Shields PE, Havel CM, Watson JA (1986) Isoprenoid synthesis in *Halobacterium halobium*. Modulation of 3-hydroxy- 3-methylglutaryl coenzyme a concentration in response to mevalonate availability. *J Biol Chem* **261:** 3578-3583.

Cann IK, Komori K, Toh H, Kanai S, Ishino Y (1998) A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc Natl Acad Sci U S A* **95:** 14250-14255

Cavicchioli R (2007) *Archaea: Molecular and cellular biology*: ASM Press.

Charlebois RL, Lam WL, Cline SW, Doolittle WF (1987) Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaebacterium. *Proc Natl Acad Sci U S A* **84:** 8530-8534

Charlebois RL, Schalkwyk LC, Hofman JD, Doolittle WF (1991) Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. *J Mol Biol* **222:** 509-524

Chen Z, Speck C, Wendel P, Tang C, Stillman B, Li H (2008) The architecture of the DNA replication origin recognition complex in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **105:** 10326-10331

Chuang RY, Kelly TJ (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci U S A* **96:** 2656-2661

Clarke L, Carbon J (1976) A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. *Cell* **9:** 91-99

Cline SW, Lam WL, Charlebois RL, Schalkwyk LC, Doolittle WF (1989) Transformation methods for halophilic archaebacteria. *Can J Microbiol* **35:** 148-152

Clyne RK, Kelly TJ (1995) Genetic analysis of an ARS element from the fission yeast *Schizosaccharomyces pombe*. *Embo J* **14:** 6348-6357

Coin F, Bergmann E, Tremeau-Bravard A, Egly JM (1999) Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. *Embo J* **18:** 1357-1366

Coin F, Oksenych V, Egly JM (2007) Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol Cell* **26:** 245-256

Costa RM, Morgante PG, Berra CM, Nakabashi M, Bruneau D, Bouchez D, Sweder KS, Van Sluys MA, Menck CF (2001) The participation of AtXPB1, the XPB/RAD25 homologue gene from *Arabidopsis thaliana*, in DNA repair and plant development. *Plant J* **28:** 385-395

Crowley DJ, Boubriak I, Berquist BR, Clark M, Richard E, Sullivan L, DasSarma S, McCready S (2006) The *uvrA*, *uvrB* and *uvrC* genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1. *Saline Systems* **2:** 11

Cvetic C, Walter JC (2005) Eukaryotic origins of DNA replication: could you please be more specific? *Semin Cell Dev Biol* **16:** 343-353

Dai J, Chuang RY, Kelly TJ (2005) DNA replication origins in the *Schizosaccharomyces pombe* genome. *Proc Natl Acad Sci U S A* **102:** 337-342

Danis E, Brodolin K, Menut S, Maiorano D, Girard-Reydet C, Mechali M (2004) Specification of a DNA replication origin by a transcription complex. *Nat Cell Biol* **6:** 721-730

Danson MJ, Hough DW (1997) The structural basis of protein halophilicity. *Comp Biochem Physiol A Physiol* **117:** 307-312

De Felice M, Esposito L, Pucci B, Carpentieri F, De Falco M, Rossi M, Pisani FM (2003) Biochemical characterization of a CDC6-like protein from the crenarchaeon *Sulfolobus solfataricus*. *J Biol Chem* **278:** 46424-46431

De Palmenaer D, Siguier P, Mahillon J (2008) IS4 family goes genomic. *BMC Evol Biol* **8:** 18

DeLong EF, Pace NR (2001) Environmental diversity of bacteria and archaea. *Syst Biol* **50:** 470-478

Dennis PP (1997) Ancient ciphers: translation in Archaea. *Cell* **89:** 1007-1010

DePamphilis ML (1999) Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* **21:** 5-16

Diffley JF, Cocker JH, Dowell SJ, Rowley A (1994) Two steps in the assembly of complexes at yeast replication origins *in vivo*. *Cell* **78:** 303-316

Dijkwel PA, Hamlin JL (1995) The Chinese hamster dihydrofolate reductase origin consists of multiple potential nascent-strand start sites. *Mol Cell Biol* **15:** 3023-3031

Dijkwel PA, Hamlin JL (1997) Mapping replication origins by neutral/neutral two-dimensional gel electrophoresis. *Methods* **13:** 235-245

Dixon NE, Kornberg A (1984) Protein HU in the enzymatic replication of the chromosomal origin of *Escherichia coli*. *Proc Natl Acad Sci U S A* **81:** 424-428

Donato JJ, Chung SC, Tye BK (2006) Genome-wide hierarchy of replication origin usage in *Saccharomyces cerevisiae*. *PLoS Genet* **2:** e141

Douziech M, Coin F, Chipoulet JM, Arai Y, Ohkuma Y, Egly JM, Coulombe B (2000) Mechanism of promoter melting by the xeroderma pigmentosum complementation group B helicase of transcription factor IIH revealed by protein-DNA photo-cross-linking. *Mol Cell Biol* **20:** 8168-8177

Duan Z (2009) Genetic Analysis of Two Structure-specific Endonucleases Hef and Fen1 in Archaeon *Haloferax volcanii*. Doctor of Philosophy Thesis, University of Nottingham,

Dueber EL, Corn JE, Bell SD, Berger JM (2007) Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* **317:** 1210-1213

Duggin IG, McCallum SA, Bell SD (2008a) Chromosome replication dynamics in the archaeon *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci U S A* **105:** 16737-16742

Duggin IG, Wake RG, Bell SD, Hill TM (2008b) The replication fork trap and termination of chromosome replication. *Mol Microbiol* **70:** 1323-1333

Dutra BE, Sutera VA, Jr., Lovett ST (2007) RecA-independent recombination is efficient but limited by exonucleases. *Proc Natl Acad Sci U S A* **104:** 216-221

Edgell DR, Doolittle WF (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* **89:** 995-998

Edgell DR, Klenk HP, Doolittle WF (1997) Gene duplications in evolution of archaeal family B DNA polymerases. *J Bacteriol* **179:** 2632-2640

Egorova K, Antranikian G (2007) Biotechnology In *Archaea: Evolution, Physiology and Molecular Biology*, Garrett K (ed), 1 edn, 26, pp 296-321. Blackwell Publishing Ltd.

Elcock AH, McCammon JA (1998) Electrostatic contributions to the stability of halophilic proteins. *J Mol Biol* **280:** 731-748

Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, Randau L, Hedlund BP, Brochier-Armanet C, Kunin V, Anderson I, Lapidus A, Goltsman E, Barry K, Koonin EV, Hugenholtz P, Kyrpides N, Wanner G, Richardson P, Keller M, Stetter KO (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci U S A* **105:** 8102-8107

Erzberger JP, Mott ML, Berger JM (2006) Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat Struct Mol Biol* **13:** 676-683

Erzberger JP, Pirruccello MM, Berger JM (2002) The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. *Embo J* **21:** 4763-4773

Evans E, Moggs JG, Hwang JR, Egly JM, Wood RD (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *Embo J* **16:** 6559-6573

Fan L, Arvai AS, Cooper PK, Iwai S, Hanaoka F, Tainer JA (2006) Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair. *Mol Cell* **22:** 27-37

Feaver WJ, Svejstrup JQ, Bardwell L, Bardwell AJ, Buratowski S, Gulyas KD, Donahue TF, Friedberg EC, Kornberg RD (1993) Dual roles of a multiprotein complex from *S. cerevisiae* in transcription and DNA repair. *Cell* **75:** 1379-1387

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2008) The Pfam protein families database. *Nucleic Acids Res* **36:** D281-288

Fletcher RJ, Bishop BE, Leon RP, Sclafani RA, Ogata CM, Chen XS (2003) The structure and function of MCM from archaeal *M. Thermoautotrophicum*. *Nat Struct Biol* **10:** 160-167

Forterre P, Elie C, Kohiyama M (1984) Aphidicolin inhibits growth and DNA synthesis in halophilic arachaebacteria. *J Bacteriol* **159:** 800-802

Fox GE, Stackebrandt E, Hespell RB, Gibson J, Maniloff J, Dyer TA, Wolfe RS, Balch WE, Tanner RS, Magrum LJ, Zablen LB, Blakemore R, Gupta R, Bonen L, Lewis BJ, Stahl DA, Luehrsen KR, Chen KN, Woese CR (1980) The phylogeny of prokaryotes. *Science* **209:** 457-463

Frank AC, Lobry JR (1999) Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* **238:** 65-77

Friedberg EC, Lehmann AR, Fuchs RP (2005) Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Mol Cell* **18:** 499-505

Friedburg (2006) *DNA Repair and Mutagenesis*, 2nd edn.

Frolow F, Harel M, Sussman JL, Mevarech M, Shoham M (1996) Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat Struct Biol* **3:** 452-458

Fujikawa N, Kurumizaka H, Nureki O, Terada T, Shirouzu M, Katayama T, Yokoyama S (2003) Structural basis of replication origin recognition by the DnaA protein. *Nucleic Acids Res* **31:** 2077-2086

Gaudier M, Schuwirth BS, Westcott SL, Wigley DB (2007) Structural basis of DNA replication origin recognition by an ORC protein. *Science* **317:** 1213-1216

Gavin KA, Hidaka M, Stillman B (1995) Conserved initiator proteins in eukaryotes. *Science* **270:** 1667-1671

Georgescu RE, O'Donnell M (2007) Structural biology. Getting DNA to unwind. *Science* **317:** 1181-1182

Gerbi SA, Bielinsky AK (1997) Replication initiation point mapping. *Methods* **13:** 271-280

Gilbert DM (2001) Making sense of eukaryotic DNA replication origins. *Science* **294:** 96-100

Gilbert DM (2004) In search of the holy replicator. *Nat Rev Mol Cell Biol* **5:** 848-855

Gilles-Gonzalez MA, Gonzalez G (2005) Heme-based sensors: defining characteristics, recent developments, and regulatory hypotheses. *J Inorg Biochem* **99:** 1-22

Gilson E, Geli V (2007) How telomeres are replicated. *Nat Rev Mol Cell Biol* **8:** 825-838

Gimenes F, Takeda KI, Fiorini A, Gouveia FS, Fernandez MA (2008) Intrinsically bent DNA in replication origins and gene promoters. *Genet Mol Res* **7:** 549-558

Goldfless SJ, Morag AS, Belisle KA, Sutera VA, Jr., Lovett ST (2006) DNA repeat rearrangements mediated by DnaK-dependent replication fork repair. *Mol Cell* **21:** 595-604

Goosen N, Moolenaar GF (2008) Repair of UV damage in bacteria. *DNA Repair (Amst)* **7:** 353-379

Grabowski B, Kelman Z (2003) Archaeal DNA replication: eukaryal proteins in a bacterial context. *Annu Rev Microbiol* **57:** 487-516

Grainge I, Gaudier M, Schuwirth BS, Westcott SL, Sandall J, Atanassova N, Wigley DB (2006) Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. *J Mol Biol* **363:** 355-369

Gregor D, Pfeifer F (2005) In vivo analyses of constitutive and regulated promoters in halophilic archaea. *Microbiology* **151:** 25-33

Grigoriev A (1998) Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res* **26:** 2286-2290

Grindley ND, Whiteson KL, Rice PA (2006) Mechanisms of site-specific recombination. *Annu Rev Biochem* **75:** 567-605

Gulyas KD, Donahue TF (1992) SSL2, a suppressor of a stem-loop mutation in the HIS4 leader encodes the yeast homolog of human ERCC-3. *Cell* **69:** 1031-1042

Guy CP, Bolt EL (2005) Archaeal Hel308 helicase targets replication forks *in vivo* and *in vitro* and unwinds lagging strands. *Nucleic Acids Res* **33:** 3678-3690

Guzder SN, Sung P, Bailly V, Prakash L, Prakash S (1994) RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. *Nature* **369:** 578-581

Haldenby S (2007) Genetic Analysis of RadB, a Paralogue of the Archaeal Rad51/RecA Homologue, RadA. Doctor of Philosophy Thesis, University of Nottingham,

Haldenby S, White MF, Allers T (2009) RecA family proteins in archaea: RadA and its cousins. *Biochem Soc Trans* **37:** 102-107

Harland RM, Laskey RA (1980) Regulated replication of DNA microinjected into eggs of Xenopus laevis. *Cell* **21:** 761-771

Heller RC, Marians KJ (2006) Replisome assembly and the direct restart of stalled replication forks. *Nat Rev Mol Cell Biol* **7:** 932-943

Henle ES, Linn S (1997) Formation, prevention, and repair of DNA damage by iron/hydrogen peroxide. *J Biol Chem* **272:** 19095-19098

Hethke C, Geerlind AC, Hausner W, de Vos WM, Thomm M (1988) A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res* **24:** 2369-2376

Hoeijmakers JH (2001) Genome maintenance mechanisms for preventing cancer. *Nature* **411:** 366-374

Hofman JD, Schalkwyk LC, Doolittle WF (1986) ISH51: a large degenerate family of insertion sequence-like elements in the genome of the archaebacterium, *Halobacterium volcanii*. *Nucleic Acids Res* **14:** 6983-7000

Holmes ML, Dyall-Smith ML (2000) Sequence and expression of a halobacterial beta-galactosidase gene. *Mol Microbiol* **36:** 114-122

Holmes ML, Nuttall SD, Dyall-Smith ML (1991) Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. *J Bacteriol* **173:** 3807-3813

Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417:** 63-67

Huberman JA (1998) DNA replication. Choosing a place to begin. *Science* **281:** 929-930

Huberman JA, Riggs AD (1968) On the mechanism of DNA replication in mammalian chromosomes. *J Mol Biol* **32:** 327-341

Huberman JA, Spotila LD, Nawotka KA, el-Assouli SM, Davis LR (1987) The *in vivo* replication origin of the yeast 2 microns plasmid. *Cell* **51:** 473-481

Huberman JA, Zhu JG, Davis LR, Newlon CS (1988) Close association of a DNA replication origin and an ARS element on chromosome III of the yeast, *Saccharomyces cerevisiae*. *Nucleic Acids Res* **16:** 6373-6384

Huth JR, Bewley CA, Nissen MS, Evans JN, Reeves R, Gronenborn AM, Clore GM (1997) The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat Struct Biol* **4:** 657-665

Hyrien O (2000) Mechanisms and consequences of replication fork arrest. *Biochimie* **82:** 5-17

Hyrien O, Maric C, Mechali M (1995) Transition in specification of embryonic metazoan DNA replication origins. *Science* **270:** 994-997

Ishimi Y (1997) A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex. *J Biol Chem* **272:** 24508-24513

Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* **146:** 11-31

Iyer VN, Szybalski W (1964) Mitomycins and Porfiromycin: Chemical Mechanism of Activation and Cross-Linking of DNA. *Science* **145:** 55-58

Jacob F, Brenner S (1963) [On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon.]. *C R Hebd Seances Acad Sci* **256:** 298-300

Johnson CH, Mori T, Xu Y (2008) A cyanobacterial circadian clockwork. *Curr Biol* **18:** R816-R825

Kalejta RF, Li X, Mesner LD, Dijkwel PA, Lin HB, Hamlin JL (1998) Distal sequences, but not *ori-beta*/OBR-1, are essential for initiation of DNA replication in the Chinese hamster DHFR origin. *Mol Cell* **2:** 797-806

Kasiviswanathan R, Shin JH, Kelman Z (2005) Interactions between the archaeal Cdc6 and MCM proteins modulate their biochemical properties. *Nucleic Acids Res* **33:** 4940-4950

Kasiviswanathan R, Shin JH, Kelman Z (2006) DNA binding by the *Methanothermobacter thermautotrophicus* Cdc6 protein is inhibited by the minichromosome maintenance helicase. *J Bacteriol* **188:** 4577-4580

Katayama T, Kubota T, Kurokawa K, Crooke E, Sekimizu K (1998) The initiator function of DnaA protein is negatively regulated by the sliding clamp of the *E. coli* chromosomal replicase. *Cell* **94:** 61-71

Kato J, Katayama T (2001) Hda, a novel DnaA-related protein, regulates the replication cycle in *Escherichia coli*. *Embo J* **20:** 4253-4262

Keeling PJ, Charlebois RL, Doolittle WF (1994) Archaebacterial genomes: eubacterial form and eukaryotic content. *Curr Opin Genet Dev* **4:** 816-822

Kelman Z, White MF (2005) Archaeal DNA replication and repair. *Curr Opin Microbiol* **8:** 669-676

Kennedy SP, Ng WV, Salzberg SL, Hood L, DasSarma S (2001) Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res* **11:** 1641-1650

Khan SA (2005) Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid* **53:** 126-136

Kinch LN, Ginalski K, Rychlewski L, Grishin NV (2005) Identification of novel restriction endonuclease-like fold families among hypothetical proteins. *Nucleic Acids Res* **33:** 3598-3605

Kitagawa R, Mitsuki H, Okazaki T, Ogawa T (1996) A novel DnaA protein-binding site at 94.7 min on the *Escherichia coli* chromosome. *Mol Microbiol* **19:** 1137-1147

Kitagawa R, Ozaki T, Moriya S, Ogawa T (1998) Negative control of replication initiation by a novel chromosomal locus exhibiting exceptional affinity for *Escherichia coli* DnaA protein. *Genes Dev* **12:** 3032-3043

Kitsberg D, Selig S, Keshet I, Cedar H (1993) Replication structure of the human beta-globin gene domain. *Nature* **366:** 588-590

Klungsoyr HK, Skarstad K (2004) Positive supercoiling is generated in the presence of *Escherichia coli* SeqA protein. *Mol Microbiol* **54:** 123-131

Kobayashi T, Rein T, DePamphilis ML (1998) Identification of primary initiation sites for DNA replication in the hamster dihydrofolate reductase gene initiation zone. *Mol Cell Biol* **18:** 3266-3277

Kogoma T (1997) Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* **61:** 212-238

Komori K, Fujikane R, Shinagawa H, Ishino Y (2002) Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes Genet Syst* **77:** 227-241

Komori K, Miyata T, Daiyasu H, Toh H, Shinagawa H, Ishino Y (2000) Domain analysis of an archaeal RadA protein for the strand exchange activity. *J Biol Chem* **275:** 33791-33797

Kong D, DePamphilis ML (2001) Site-specific DNA binding of the *Schizosaccharomyces pombe* origin recognition complex is determined by the Orc4 subunit. *Mol Cell Biol* **21:** 8095-8103

Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36:** 6688-6719

Kornberg B (1992) *DNA Replication*, 2nd edn.: W. H. Freeman and Company.

Krysan PJ, Calos MP (1991) Replication initiates at multiple locations on an autonomously replicating plasmid in human cells. *Mol Cell Biol* **11:** 1464-1472

Krysan PJ, Haase SB, Calos MP (1989) Isolation of human sequences that replicate autonomously in human cells. *Mol Cell Biol* **9:** 1026-1033

Labib K, Diffley JF (2001) Is the MCM2-7 complex the eukaryotic DNA replication fork helicase? *Curr Opin Genet Dev* **11:** 64-70

Labib K, Gambus A (2007) A key role for the GINS complex at DNA replication forks. *Trends Cell Biol* **17:** 271-278

Lambert S, Carr AM (2005) Checkpoint responses to replication fork barriers. *Biochimie* **87:** 591-602

Lambert S, Froget B, Carr AM (2007) Arrested replication fork processing: interplay between checkpoints and recombination. *DNA Repair (Amst)* **6:** 1042-1061

Large A, Stamme C, Lange C, Duan Z, Allers T, Soppa J, Lund PA (2007) Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential *cct1* gene. *Mol Microbiol* **66:** 1092-1106

Lee DG, Bell SP (1997) Architecture of the yeast origin recognition complex bound to origins of DNA replication. *Mol Cell Biol* **17:** 7159-7168

Lee JK, Moon KY, Jiang Y, Hurwitz J (2001) The Schizosaccharomyces pombe origin recognition complex interacts with multiple AT-rich regions of the replication origin DNA by means of the AT-hook domains of the spOrc4 protein. *Proc Natl Acad Sci U S A* **98:** 13589-13594

Lehmann AR (2003) DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy. *Biochimie* **85:** 1101-1111

Lin YC, Choi WS, Gralla JD (2005) TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. *Nat Struct Mol Biol* **12:** 603-607

Lindahl T, Wood RD (1999) Quality control by DNA repair. *Science* **286:** 1897-1905

Liu J, Smith CL, DeRyckere D, DeAngelis K, Martin GS, Berger JM (2000) Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol Cell* **6:** 637-648

Liu LF, Wang JC (1987) Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A* **84:** 7024-7027

Llorente B, Smith CE, Symington LS (2008) Break-induced replication: what is it and what is it for? *Cell Cycle* **7:** 859-864

Londei P (2005) Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* **29:** 185-200

Lopes M, Cotta-Ramusino C, Pellicioli A, Liberi G, Plevani P, Muzi-Falconi M, Newlon CS, Foiani M (2001) The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* **412:** 557-561

Lopez P, Philippe H, Myllykallio H, Forterre P (1999) Identification of putative chromosomal origins of replication in Archaea. *Mol Microbiol* **32:** 883-886

Lovett ST, Hurley RL, Sutera VA, Jr., Aubuchon RH, Lebedeva MA (2002) Crossing over between regions of limited homology in *Escherichia coli.* RecA-dependent and RecA-independent pathways. *Genetics* **160:** 851-859

Lu M, Campbell JL, Boye E, Kleckner N (1994) SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* **77:** 413-426

Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R (2004) Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* **101:** 7046-7051

MacAlpine DM, Rodriguez HK, Bell SP (2004) Coordination of replication and transcription along a Drosophila chromosome. *Genes Dev* **18:** 3094-3105

Mackiewicz P, Zakrzewska-Czerwinska J, Zawilak A, Dudek MR, Cebrat S (2004) Where does bacterial replication start? Rules for predicting the *oriC* region. *Nucleic Acids Res* **32:** 3781-3791

Maisnier-Patin S, Malandrin L, Birkeland NK, Bernander R (2002) Chromosome replication patterns in the hyperthermophilic euryarchaea *Archaeoglobus fulgidus* and *Methanocaldococcus* (*Methanococcus*) *jannaschii*. *Mol Microbiol* **45:** 1443-1450

Majernik AI, Chong JP (2008) A conserved mechanism for replication origin recognition and binding in archaea. *Biochem J* **409:** 511-518

Majernik AI, Jenkinson ER, Chong JP (2004) DNA replication in thermophiles. *Biochem Soc Trans* **32:** 236-239

Makarova KS, Koonin EV (2003) Comparative genomics of Archaea: how much have we learned in six years, and what's next? *Genome Biol* **4:** 115

Makarova KS, Sorokin AV, Novichkov PS, Wolf YI, Koonin EV (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol Direct* **2:** 33

Marahrens Y, Stillman B (1992) A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255:** 817-823

Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, Ke Z, Krylov D, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Thanki N, Yamashita RA, Yin JJ, Zhang D, Bryant SH (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35:** D237-240

Masai H, Asai T, Kubota Y, Arai K, Kogoma T (1994) *Escherichia coli* PriA protein is essential for inducible and constitutive stable DNA replication. *Embo J* **13:** 5338-5345

Matsunaga F, Forterre P, Ishino Y, Myllykallio H (2001) *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proc Natl Acad Sci U S A* **98:** 11152-11157

Matsunaga F, Glatigny A, Mucchielli-Giorgi MH, Agier N, Delacroix H, Marisa L, Durosay P, Ishino Y, Aggerbeck L, Forterre P (2007) Genomewide and biochemical analyses of DNA-binding activity of Cdc6/Orc1 and Mcm proteins in Pyrococcus sp. *Nucleic Acids Res* **35:** 3214-3222

Matsunaga F, Norais C, Forterre P, Myllykallio H (2003) Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep* **4:** 154-158

Mayr E (1998) Two empires or three? *Proc Natl Acad Sci U S A* **95:** 9720-9723

McCready S (1996) The repair of ultraviolet light-induced DNA damage in the halophilic archaebacteria, *Halobacterium cutirubrum*, *Halobacterium halobium* and *Haloferax volcanii*. *Mutat Res* **364:** 25-32

McCready S, Marcello L (2003) Repair of UV damage in *Halobacterium salinarum*. *Biochem Soc Trans* **31:** 694-698

McGlynn P, Lloyd RG (2002) Recombinational repair and restart of damaged replication forks. *Nat Rev Mol Cell Biol* **3:** 859-870

Mechali M, Kearsey S (1984) Lack of specific sequence requirement for DNA replication in Xenopus eggs compared with high sequence specificity in yeast. *Cell* **38:** 55-64

Messer W (2002) The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiol Rev* **26:** 355-374

Messer W, Blaesing F, Majka J, Nardmann J, Schaper S, Schmidt A, Seitz H, Speck C, Tungler D, Wegrzyn G, Weigel C, Welzeck M, Zakrzewska-Czerwinska J (1999) Functional domains of DnaA proteins. *Biochimie* **81:** 819-825

Messer W, Weigel C (1997) DnaA initiator--also a transcription factor. *Mol Microbiol* **24:** 1-6

Mizushima T, Takahashi N, Stillman B (2000) Cdc6p modulates the structure and DNA binding activity of the origin recognition complex *in vitro*. *Genes Dev* **14:** 1631-1641

Monson EK, Weinstein M, Ditta GS, Helinski DR (1992) The FixL protein of Rhizobium meliloti can be separated into a heme-binding oxygen-sensing domain and a functional C-terminal kinase domain. *Proc Natl Acad Sci U S A* **89:** 4280-4284

Morgante PG, Berra CM, Nakabashi M, Costa RM, Menck CF, Van Sluys MA (2005) Functional XPB/RAD25 redundancy in Arabidopsis genome: characterization of AtXPB2 and expression analysis. *Gene* **344:** 93-103

Mulcair MD, Schaeffer PM, Oakley AJ, Cross HF, Neylon C, Hill TM, Dixon NE (2006) A molecular mousetrap determines polarity of termination of DNA replication in E. coli. *Cell* **125:** 1309-1319

Mullakhanbhai MF, Larsen H (1975) *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch Microbiol* **104:** 207-214

Myllykallio H, Lopez P, Lopez-Garcia P, Heilig R, Saurin W, Zivanovic Y, Philippe H, Forterre P (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* **288:** 2212-2215

Naumovski L, Friedberg EC (1983) A DNA repair gene required for the incision of damaged DNA is essential for viability in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **80:** 4818-4821

Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9:** 27-43

Newlon CS, Theis JF (1993) The structure and function of yeast ARS elements. *Curr Opin Genet Dev* **3:** 752-758

Ng WL, DasSarma S (1993) Minimal replication origin of the 200-kilobase *Halobacterium* plasmid pNRC100. *J Bacteriol* **175:** 4584-4596

Nguyen VQ, Co C, Li JJ (2001) Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. *Nature* **411:** 1068-1073

Nieduszynski CA, Knox Y, Donaldson AD (2006) Genome-wide identification of replication origins in yeast by comparative genomics. *Genes Dev* **20:** 1874-1879

Noll DM, Mason TM, Miller PS (2006) Formation and repair of interstrand cross-links in DNA. *Chem Rev* **106:** 277-301

Norais C (2007) Etude de la Réplication de l'ADN chez l'Archaea Halophile *Haloferax volcanii*. Docteur en Sciences Thesis, Institut de Génetique et Microbiologie, Université Paris-Sud 11, Paris

Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet* **3:** e77

Ogrunc M, Becker DF, Ragsdale SW, Sancar A (1998) Nucleotide excision repair in the third kingdom. *J Bacteriol* **180:** 5796-5798

Okuno Y, Satoh H, Sekiguchi M, Masukata H (1999) Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. *Mol Cell Biol* **19:** 6699-6709

Ortenberg R, Rozenblatt-Rosen O, Mevarech M (2000) The extremely halophilic archaeon *Haloferax volcanii* has two very different dihydrofolate reductases. *Mol Microbiol* **35:** 1493-1505

Park E, Guzder SN, Koken MH, Jaspers-Dekker I, Weeda G, Hoeijmakers JH, Prakash S, Prakash L (1992) RAD25 (SSL2), the yeast homolog of the human xeroderma pigmentosum group B DNA repair gene, is essential for viability. *Proc Natl Acad Sci U S A* **89:** 11416-11420

Paul S, Bag SK, Das S, Harvill ET, Dutta C (2008) Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. *Genome Biol* **9:** R70

Prakash S, Prakash L (2000) Nucleotide excision repair in yeast. *Mutat Res* **451:** 13-24

Qiu H, Park E, Prakash L, Prakash S (1993) The *Saccharomyces cerevisiae* DNA repair gene RAD25 is required for transcription by RNA polymerase II. *Genes Dev* **7:** 2161-2171

Rao H, Stillman B (1995) The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. *Proc Natl Acad Sci U S A* **92:** 2224-2228

Remus D, Beall EL, Botchan MR (2004) DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *Embo J* **23:** 897-907

Richards JD, Cubeddu L, Roberts J, Liu H, White MF (2008) The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner. *J Mol Biol* **376:** 634-644

Robinson NP, Bell SD (2005) Origins of DNA replication in the three domains of life. *Febs J* **272:** 3757-3766

Robinson NP, Bell SD (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci U S A* **104:** 5806-5811

Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. *Cell* **116:** 25-38

Robinson NP, Blood KA, McCallum SA, Edwards PAW, Bell SD (2004) Sister chromatid junctions in the hyperthermophilic archaeon *Sulfolobus solfataricus EMBO J* **26:** 816–824

Roten CA, Gamba P, Barblan JL, Karamata D (2002) Comparative Genometrics (CG): a database dedicated to biometric comparisons of whole genomes. *Nucleic Acids Res* **30:** 142-144

Rother M, Metcalf WW (2005) Genetic technologies for Archaea. *Curr Opin Microbiol* **8:** 745-751

Rowley A, Cocker JH, Harwood J, Diffley JF (1995) Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. *Embo J* **14:** 2631-2641

Rudolph CJ, Upton AL, Lloyd RG (2007) Replication fork stalling and cell cycle arrest in UV-irradiated *Escherichia coli*. *Genes Dev* **21:** 668-681

Saha S, Shan Y, Mesner LD, Hamlin JL (2004) The promoter of the Chinese hamster ovary dihydrofolate reductase gene regulates the activity of the local origin and helps define its boundaries. *Genes Dev* **18:** 397-410

Salerno V, Napoli A, White MF, Rossi M, Ciaramella M (2003) Transcriptional response to DNA damage in the archaeon *Sulfolobus solfataricus*. *Nucleic Acids Res* **31:** 6127-6138

Sandler SJ, Satin LH, Samra HS, Clark AJ (1996) *recA*-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res* **24:** 2125-2132

Sasaki T, Gilbert DM (2007) The many faces of the origin recognition complex. *Curr Opin Cell Biol* **19:** 337-343

Sasaki T, Ramanathan S, Okuno Y, Kumagai C, Shaikh SS, Gilbert DM (2006) The Chinese hamster dihydrofolate reductase replication origin decision point follows activation of transcription and suppresses initiation of replication within transcription units. *Mol Cell Biol* **26:** 1051-1062

Sasaki T, Sawado T, Yamaguchi M, Shinomiya T (1999) Specification of regions of DNA replication initiation during embryogenesis in the 65-kilobase DNApolalpha-dE2F locus of *Drosophila melanogaster*. *Mol Cell Biol* **19:** 547-555

Schaeffer L, Roy R, Humbert S, Moncollin V, Vermeulen W, Hoeijmakers JH, Chambon P, Egly JM (1993) DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* **260:** 58-63

Schaper S, Messer W (1995) Interaction of the initiator protein DnaA of Escherichia coli with its DNA target. *J Biol Chem* **270:** 17622-17626

Sclafani RA, Holzen TM (2007) Cell cycle regulation of DNA replication. *Annu Rev Genet* **41:** 237-280

Sernova NV, Gelfand MS (2008) Identification of replication origins in prokaryotic genomes. *Brief Bioinform* **9:** 376-391

Shakibai N, Ishidate K, Reshetnyak E, Gunji S, Kohiyama M, Rothfield L (1998) High-affinity binding of hemimethylated oriC by Escherichia coli membranes is mediated by a multiprotein system that includes SeqA and a newly identified factor, SeqB. *Proc Natl Acad Sci U S A* **95:** 11117-11121

Shin JH, Heo GY, Kelman Z (2009) The *Methanothermobacter thermautotrophicus* MCM Helicase Is Active as a Hexameric Ring. *J Biol Chem* **284:** 540-546

Singleton MR, Morales R, Grainge I, Cook N, Isupov MN, Wigley DB (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. *J Mol Biol* **343:** 547-557

Slupphaug G, Kavli B, Krokan HE (2003) The interacting pathways for prevention and repair of oxidative DNA damage. *Mutat Res* **531:** 231-251

Smith DW, Garland AM, Herman G, Enns RE, Baker TA, Zyskind JW (1985) Importance of state of methylation of *oriC* GATC sites in initiation of DNA replication in Escherichia coli. *Embo J* **4:** 1319-1326

Soppa J (2006) From genomes to function: haloarchaea as model organisms. *Microbiology* **152:** 585-590

Speck C, Messer W (2001) Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *Embo J* **20:** 1469-1476

Speck C, Stillman B (2007) Cdc6 ATPase activity regulates ORC x Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. *J Biol Chem* **282:** 11705-11714

Speck C, Weigel C, Messer W (1997) From footprint to toeprint: a close-up of the DnaA box, the binding site for the bacterial initiator protein DnaA. *Nucleic Acids Res* **25:** 3242-3247

Speck C, Weigel C, Messer W (1999) ATP- and ADP-dnaA protein, a molecular switch in gene regulation. *Embo J* **18:** 6169-6176

Stinchcomb DT, Struhl K, Davis RW (1979) Isolation and characterisation of a yeast chromosomal replicator. *Nature* **282:** 39-43

Struhl K, Stinchcomb DT, Scherer S, Davis RW (1979) High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. *Proc Natl Acad Sci U S A* **76:** 1035-1039

Sung P, Guzder SN, Prakash L, Prakash S (1996) Reconstitution of TFIIH and requirement of its DNA helicase subunits, Rad3 and Rad25, in the incision step of nucleotide excision repair. *J Biol Chem* **271:** 10821-10826

Sung P, Higgins D, Prakash L, Prakash S (1988) Mutation of lysine-48 to arginine in the yeast RAD3 protein abolishes its ATPase and DNA helicase activities but not the ability to bind ATP. *Embo J* **7:** 3263-3269

Sung P, Klein H (2006) Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol* **7:** 739-750

Sutton MD, Kaguni JM (1997) The *Escherichia coli dnaA* gene: four functional domains. *J Mol Biol* **274:** 546-561

Tada S, Kundu LR, Enomoto T (2008) Insight into initiator-DNA interactions: a lesson from the archaeal ORC. *Bioessays* **30:** 208-211

Takayama Y, Kamimura Y, Okawa M, Muramatsu S, Sugino A, Araki H (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev* **17:** 1153-1165

Taylor JH (1968) Rates of chain growth and units of replication in DNA of mammalian chromosomes. *J Mol Biol* **31:** 579-594

Thomm H (2007) Transcriptional mechanisms. In *Archaea: Evolution, Physiology and Molecular Biology*, Garrett K (ed), 1 edn, 16, pp 185-197. Blackwell Publishing Ltd

Tye BK (1999) MCM proteins in DNA replication. *Annu Rev Biochem* **68:** 649-686

van de Vossenberg JL, Driessen AJ, Konings WN (1998) The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles* **2:** 163-170

Van Houten B, Croteau DL, DellaVecchia MJ, Wang H, Kisker C (2005) 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutat Res* **577:** 92-117

Vas A, Leatherwood J (2000) Where does DNA replication start in archaea? *Genome Biol* **1:** REVIEWS1020

Vaughn JP, Dijkwel PA, Hamlin JL (1990) Replication initiates in a broad zone in the amplified CHO dihydrofolate reductase domain. *Cell* **61:** 1075-1087

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-Tillson H, Pfannkoch C, Rogers YH, Smith HO (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66-74

von Freiesleben U, Rasmussen KV, Schaechter M (1994) SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. *Mol Microbiol* **14:** 763-772

Wang L, Lin CM, Brooks S, Cimbora D, Groudine M, Aladjem MI (2004) The human beta-globin replication initiation region consists of two modular independent replicators. *Mol Cell Biol* **24:** 3373-3386

Wang Z, Svejstrup JQ, Feaver WJ, Wu X, Kornberg RD, Friedberg EC (1994) Transcription factor b (TFIIH) is required during nucleotide-excision repair in yeast. *Nature* **368:** 74-76

Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171:** 737-738

Weeda G, van Ham RC, Masurel R, Westerveld A, Odijk H, de Wit J, Bootsma D, van der Eb AJ, Hoeijmakers JH (1990a) Molecular cloning and biological characterization of the human excision repair gene ERCC-3. *Mol Cell Biol* **10:** 2570-2581

Weeda G, van Ham RC, Vermeulen W, Bootsma D, van der Eb AJ, Hoeijmakers JH (1990b) A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell* **62:** 777-791

Wendoloski D, Ferrer C, Dyall-Smith ML (2001) A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. *Microbiology* **147:** 959-964.

Werner F (2007) Structure and function of archaeal RNA polymerases. *Mol Microbiol* **65:** 1395-1404

White MF (2003) Archaeal DNA repair: paradigms and puzzles. *Biochem Soc Trans* **31:** 690-693

White MF, Bell SD (2002) Holding it together: chromatin in the Archaea. *Trends Genet* **18:** 621-626

Wigley DB (2009) ORC proteins: marking the start. *Curr Opin Struct Biol*

Woese CR (1987) Bacterial evolution. *Microbiol Rev* **51:** 221-271

Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74:** 5088-5090

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87:** 4576-4579

Woods WG, Dyall-Smith ML (1997) Construction and analysis of a recombination-deficient *(radA*) mutant of *Haloferax volcanii*. *Mol Microbiol* **23:** 791-797

Wyatt MD, Pittman DL (2006) Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem Res Toxicol* **19:** 1580-1594

Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* **294:** 2357-2360

Zeikus JG, Wolfe RS (1972) Methanobacterium thermoautotrophicus sp. n., an anaerobic, autotrophic, extreme thermophile. *J Bacteriol* **109:** 707-715

Zhang R, Zhang CT (1994) Z curves, an intutive tool for visualizing and analyzing the DNA sequences. *J Biomol Struct Dyn* **11:** 767-782

Zhang R, Zhang CT (2003) Multiple replication origins of the archaeon *Halobacterium* species NRC-1. *Biochem Biophys Res Commun* **302:** 728-734

Zhang R, Zhang CT (2005) Identification of replication origins in archaeal genomes based on the Z-curve method. *Archaea* **1:** 335-346

Zhao A, Gray FC, MacNeill SA (2006) ATP- and NAD+-dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. *Mol Microbiol* **59:** 743-752

Zuckerkandl E, Pauling L (1965) Molecules as documents of evolutionary history. *J Theor Biol* **8:** 357-366

Chapter 7: Appendices

7.1 COG1474 consensus sequence

Cdc6-related protein, AAA superfamily ATPase consensus sequence

ETIFKNKDVLLEDYIPEELPHREEEINQLASFLAPALRGERPSNIIIYGPTGTGKTA
TVKFVMEELEESSANVEVVYINCLELRTPYQVLSKILNKLGKVPLTGDSSLEILKRL
YDNLSKKGKTVIVILDEVDALVDKDGEVLYSLLRAPGENKVKVSIIAVSNDDKFLDY
LDPRVKSSLGPSEIVFPPYTAEELYDILRERVEEGFSAGVIDDDVLKLIAALVAAES
GDARKAIDILRRAGEIAEREGSRKVSEDHVREAQEEIERDVLEEVLKTLPLHQKIVL
LAIVELTVEISTGELYDVYESLCERLRTSQRRFSDIISELEGLGIVSASLISRGERG
RTREISLDLDPEVIREILKLDLRL

Consensus sequence obtained from the NCBI conserved core domain database, January 2009.

7.2 *bga* translocation from pHV4 to the main chromosome

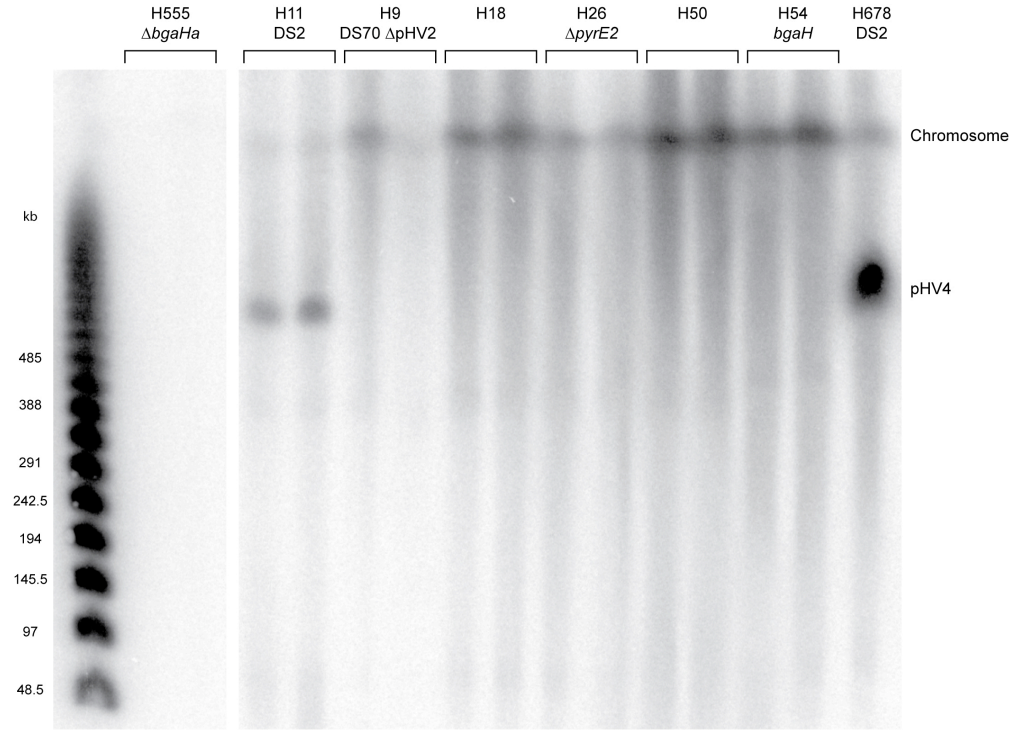


Figure 7.1: *bga* replicon location in *Haloferax volcanii*

Southern blot of intact *H. volcanii* DNA from various strains displayed on a PFG and probed with *bga* (pTA506 *Bam*HI/*Eco*RV). Excluding H555 and H678, each strain is derived from the strain to its left. H555 (Δ*pyrE2*, Δ*bgaHv*), H11 (DS2), H9 (DS70), H18 (p*yrE2*+::[Δ*pyrE2* Nov^{R}]), H26 (Δ*pyrE2*), H50 (Δ*pyrE2*, bgaHv::[*bgaH pyrE2*+]), H54 (Δ*pyrE2*, *bgaHa*), H678 (DS2).

7.3 COG1061 consensus sequence

DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair] consensus sequence

YGDLKQYLSSKGAEELADYVLDEGLPLKLIVAFEFELRPYQEEALDALVKNRRTERR
GVIVLPTGAGKTVVAAEAIAELKRSTLVLVPTKELLDQWAEALKKFLLLNDEIGIYG
GGEKELEPAKVTVATVQTLARRQLLDEFLGNEFGLIIFDEVHHLPAPSYRRILELLS
AAYPRLGLTATPEREDGGRIGDLFDLIGPIVYEVSLKELIDEGYLAPYKYVEIKVTL
TEDEEREYAKESARFRELLRARGTLRAENEARRIAIASERKIAAVRGLLLKHARGDK
TLIFASDVEHAYEIAKLFLAPGIVEAITGETPKEEREAILERFRTGGIKVLVTVKVL
DEGVDIPDADVLIILRPTGSRRLFIQRLGRGLRPAEGKEDTLALDYSLVPDDLGEED
IARRRRLFLIRKGYTYRLLTADEEGELIPNLILGIKGYRDALL

VEVVKKRCQEIDVLEEYDFRNDHRNPDLDILKPSTQIRPYQEKSLSKMFGNGRARSG
IIVLPCGAGKTLVGITAACTIKKSVIVLCTSSVSVMQWRQQFLQWCTLQENCAVFTS
DNKEMFQSGLVVSTYSMVANTRVMDFLTGREWGFIILDEVHVVPAAMFRRVVSTIAA
HA-KLGLTATLVRED-DKIGDLNFLIGPKLYEANWMELSQKGHIANVQCAEVWCPMT
AEFYQEYLRETARKR--------------MLLYIMNPTKFQACQFLIQYHERGDKII
VFSDNVYALQEYALKMGKP----FIYGSTPQQERMNILQNFQYNDINTIFLSKVGDT
SIDLPEATCLIQISSYGSRRQEAQRLGRILRARRNDEGFNAFFYSLVSKDTQEMYYS
TKRQAFLVDQGYAFKVITHLHGMENIPNLAYASPRERRELL

Consensus sequence for the two domains of COG1061, obtained from the NCBI conserved core domain database, February 2009.

7.4 Genetic and Physical Mapping of DNA Replication Origins in *Haloferax volcanii*

Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet* 3: e77

OPEN ACCESS Freely available online PLoS GENETICS

Genetic and Physical Mapping of DNA Replication Origins in *Haloferax volcanii*

Cédric Norais[1,2☯], Michelle Hawkins[3☯], Amber L. Hartman[4¤], Jonathan A. Eisen[5¤], Hannu Myllykallio[1,2*], Thorsten Allers[3*]

1 Institut de Génétique et Microbiologie, Université Paris-Sud, Orsay, France, **2** CNRS, UMR8621, Orsay, France, **3** Institute of Genetics, University of Nottingham, Nottingham, United Kingdom, **4** Johns Hopkins University, Baltimore, Maryland, United States of America, **5** The Institute for Genomic Research, Rockville, Maryland, United States of America

The halophilic archaeon *Haloferax volcanii* has a multireplicon genome, consisting of a main chromosome, three secondary chromosomes, and a plasmid. Genes for the initiator protein Cdc6/Orc1, which are commonly located adjacent to archaeal origins of DNA replication, are found on all replicons except plasmid pHV2. However, prediction of DNA replication origins in *H. volcanii* is complicated by the fact that this species has no less than 14 *cdc6/orc1* genes. We have used a combination of genetic, biochemical, and bioinformatic approaches to map DNA replication origins in *H. volcanii*. Five autonomously replicating sequences were found adjacent to *cdc6/orc1* genes and replication initiation point mapping was used to confirm that these sequences function as bidirectional DNA replication origins in vivo. Pulsed field gel analyses revealed that *cdc6/orc1*-associated replication origins are distributed not only on the main chromosome (2.9 Mb) but also on pHV1 (86 kb), pHV3 (442 kb), and pHV4 (690 kb) replicons. Gene inactivation studies indicate that linkage of the initiator gene to the origin is not required for replication initiation, and genetic tests with autonomously replicating plasmids suggest that the origin located on pHV1 and pHV4 may be dominant to the principal chromosomal origin. The replication origins we have identified appear to show a functional hierarchy or differential usage, which might reflect the different replication requirements of their respective chromosomes. We propose that duplication of *H. volcanii* replication origins was a prerequisite for the multireplicon structure of this genome, and that this might provide a means for chromosome-specific replication control under certain growth conditions. Our observations also suggest that *H. volcanii* is an ideal organism for studying how replication of four replicons is regulated in the context of the archaeal cell cycle.

Citation: Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, et al. (2007) Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. PLoS Genet 3(5): e77. doi:10.1371/journal.pgen.0030077

Introduction

In all prokaryotic organisms, and in certain unicellular eukaryotes, DNA replication is thought to initiate at well-defined chromosomal sites. These origins of replication serve as assembly sites for the protein machinery that unwinds the DNA duplex and initiates bidirectional DNA synthesis [1,2]. Bacterial chromosomes typically carry a single replication origin *(oriC)* and cognate initiator protein (e.g., DnaA), whereas eukaryotic chromosomes contain large numbers of replication origins, which are bound by a multiprotein origin recognition complex (ORC). Archaea use DNA replication proteins similar to those of eukaryotes but have circular chromosomes like bacteria [3]. Relatively little is known about origin utilization in archaea, and the available data suggest major differences in how chromosomes are replicated in the key archaeal groups. In particular, the chromosome of *Pyrococcus abyssi* (Euryarchaeota) is replicated from a single *oriC* [4], whereas three different *oriC*s are used to replicate the single chromosome of *Sulfolobus* species (Crenarchaeota) [5,6].

Archaeal replication origins consist of a long intergenic sequence containing an A/T-rich duplex unwinding element (DUE), which facilitates the local duplex opening required for replication fork assembly. The intergenic region is typically located upstream of a *cdc6/orc1* gene, which encodes a putative initiator protein that is homologous to both a subunit (Orc1) of eukaryotic ORC and the helicase loader Cdc6. Protein complexes formed by archaeal initiator proteins could therefore have a dual function in origin recognition and loading of minichromosome maintenance (MCM) helicase at the origin. However, available biochemical data are consistent with a role in origin recognition only. The intergenic region of the replication origin also carries multiple conserved sequence elements (origin recognition boxes, ORB) that are bound by Cdc6/Orc1 initiator proteins [6,7]. Binding of initiator proteins at the archaeal origin has been shown to proceed in a cooperative manner [8], suggesting that in archaea a defined multimeric initiator protein

Editor: Stephen D. Bell, Hutchison/Medical Research Council Research Centre, United Kingdom

Received December 1, 2006; **Accepted** March 5, 2007; **Published** May 18, 2007

A previous version of this article appeared as an Early Online Release on March 5, 2007 (doi:10.1371/journal.pgen.0030077.eor).

Abbreviations: 5-FOA, 5-fluoroorotic acid; ARS, autonomously replicating sequence; DUE, duplex unwinding element; MCM, minichromosome maintenance; ORB, origin recognition box; ORC, origin recognition complex; PFG, pulsed field gel; RIP, replication initiation point

* To whom correspondence should be addressed. E-mail: hannu.myllykallio@igmors.u-psud.fr (HM); thorsten.allers@nottingham.ac.uk (TA)

☯ These authors contributed equally to this work.

¤ Current address: Departments of Evolution and Ecology and Medical Microbiology and Immunology, The Genome Center, University of California Davis, Davis, California, United States of America

Author Summary

Haloferax volcanii is a member of the archaea, which are renowned for thriving in extreme environments. Archaea have circular chromosomes like bacteria but use enzymes similar to those found in eukaryotes to replicate their DNA. Few archaeal species have systems for genetics, and this has limited our understanding of DNA replication. We used genetics to map the chromosomal sites (origins) at which DNA replication initiates in *H. volcanii*. This species has a multipart genome comprising one main chromosome, three secondary chromosomes, and a plasmid. Five DNA replication origins were found and confirmed to function in vivo. All are adjacent to genes for the initiator protein Cdc6/Orc1, a common feature of archaeal replication origins. Two of the sequences are located on the main chromosome, confirming that multiple origins are often used to replicate circular chromosomes in archaea. Intriguingly, one of the origins from a secondary chromosome appears "dominant" to the principal chromosomal origin, suggesting either a hierarchy or differential usage of origins. This might reflect the different replication requirements of their respective chromosomes. Given the ease of genetic manipulation, *H. volcanii* holds great promise for studying how replication of four chromosomes is regulated in the context of the archaeal cell cycle.

complex forms at the origin. Direct support for this idea has come from recent experiments indicating that higher-order assembly of *Aeropyrum pernix* Orc proteins results in structural and topological changes in the origin DNA [9].

Current knowledge of archaeal DNA replication is based on biochemical observations, and genetic studies are needed to test the extant models. In this respect, haloarchaea are ideal model organisms since they are easily cultured and amenable to genetic manipulation [10,11]. The genome sequences of *Halobacterium* sp. NRC-1 and *Haloarcula marismortui* have revealed that haloarchaea contain multiple replicons, each with essential genes [12,13]. However, the mechanisms that coordinate the replication of these multiple replicons remain unknown.

The genome of *Haloferax volcanii* has a multireplicon structure consisting of a main chromosome of 2.9 Mb and four smaller replicons (pHV1 [86 kb], pHV2 [6.4 kb], pHV3 [442 kb], and pHV4 [690 kb]) [14]. Like other haloarchaea, the *H. volcanii* genome encodes numerous putative *cdc6/orc1* genes that are distributed amongst the different replicons (except pHV2) (Table 1). This suggests that lineage-specific duplication of replication origins and/or initiator proteins has driven dynamic genome evolution in haloarchaea. However, previous experimental studies have identified only one likely replication origin on the main chromosome of the related haloarchaeon *Halobacterium* sp. NRC-1 [15].

Analysis of the genome sequence of *H. volcanii* (Hartman et al., unpublished data) suggested the presence of numerous replication origins, prompting us to search for autonomously replicating sequence (ARS) elements corresponding to each replicon. Replication initiation point (RIP) mapping was used to confirm that these ARS elements are functional in their chromosomal context. Our genetic data suggest a genome-wide hierarchy of some ARS elements, raising the possibility of chromosome-specific origin regulation in halophilic archaea. This study also provides a framework for investigating how the coordinated replication of four replicons is achieved in halophilic archaea.

Table 1. *H. volcanii* Cdc6/Orc1 Homologues: Result of BLAST Search with Cdc6/Orc1 Consensus COG1474.1 on *H. volcanii* Genome Sequence

Gene	Episome	Contig Number	Contig Coordinates	BLAST Score	E Value	pI
orc1	Chr	455	258-1952	315	2×10^{-87}	3.98
orc2	Chr	455	570429–569203	230	7×10^{-62}	4.27
orc3	pHV4[a]	452	152-1402	218	4×10^{-58}	4.58
orc4	Chr	455	1888354–1889583	216	2×10^{-57}	4.94
orc5	Chr	455	1594707–1595912	213	9×10^{-57}	4.79
orc6	pHV3	453	201-1433	207	9×10^{-55}	5.29
orc7	pHV4	452	258668–257391	200	1×10^{-52}	4.56
orc8	pHV1	454	56369–57616	177	1×10^{-45}	4.65
orc9	Chr	455	174892–176016	172	2×10^{-44}	4.57
orc10	pHV1[b]	454	101-1327	155	3×10^{-39}	4.56
orc11	Chr	455	2162076–2162873	154	9×10^{-39}	4.61
orc12	pHV4	452	66289–65261	87	1×10^{-18}	4.69
orc13	pHV4	452	56897–55455	67	1×10^{-12}	4.51
orc14	Chr	455	2161873–2160851	67	1×10^{-12}	5.59

Origin-associated *cdc6/orc1* genes are bolded.
[a]Experimentally determined as present on chromosome, see Figure 6F.
[b]Experimentally determined as present on both pHV1 and pHV4, see Figure 1C.
doi:10.1371/journal.pgen.0030077.t001

Results

Almost all origins of DNA replication identified in archaea to date are adjacent to genes coding for a homologue of eukaryotic Cdc6 and Orc1 proteins [16]. In order to identify putative replication origins in *H. volcanii*, we performed a TBLASTN search of the genome sequence of *H. volcanii* strain DS2 (http://www.tigr.org), using the consensus Cdc6/Orc1 sequence (COG1474.1) as a query. A total of 14 potential *cdc6/orc1* genes were found on the five replicons (Table 1), although one *(orc11)* is missing sequences coding for the C-terminal winged-helix domain that is required for DNA binding [17]. The large number of *cdc6/orc1* genes found in the *H. volcanii* genome parallels the situation in other halophilic archaea; ten *cdc6/orc1* genes are found in *Halobacterium* sp. NRC-1 (three replicons) and 17 are present in *H. marismortui* (nine replicons) [12,15]. Three of the ten putative origins in *Halobacterium* sp. NRC-1 were examined in a previous study, and only one (associated with *orc7*) proved capable of supporting autonomous replication of a plasmid [15]. Thus, it is unlikely that all the *cdc6/orc1* genes in *H. volcanii* are adjacent to active origins of DNA replication.

Genetic Screen for DNA Replication Origins

To avoid any a priori bias regarding the location of replication origins, we carried out a genetic screen for ARS elements from *H. volcanii*. A partial HpaII digest was performed with genomic DNA from *H. volcanii* strain WR340 [18], which like all laboratory strains of *H. volcanii* lacks the smallest replicon pHV2 [19,20]. DNA fragments of between 4 and 8 kb were cloned in the nonreplicating *pyrE2*-marked plasmid pTA131 [21] and used to transform the *ΔpyrE2 ΔradA* strain H49 to prototrophy for uracil (*pyrE2+*). *H. volcanii radA* mutants such as H49 are defective in homologous recombination [22]; this precaution was taken to prevent integration of the plasmid into the genome by homologous recombination.

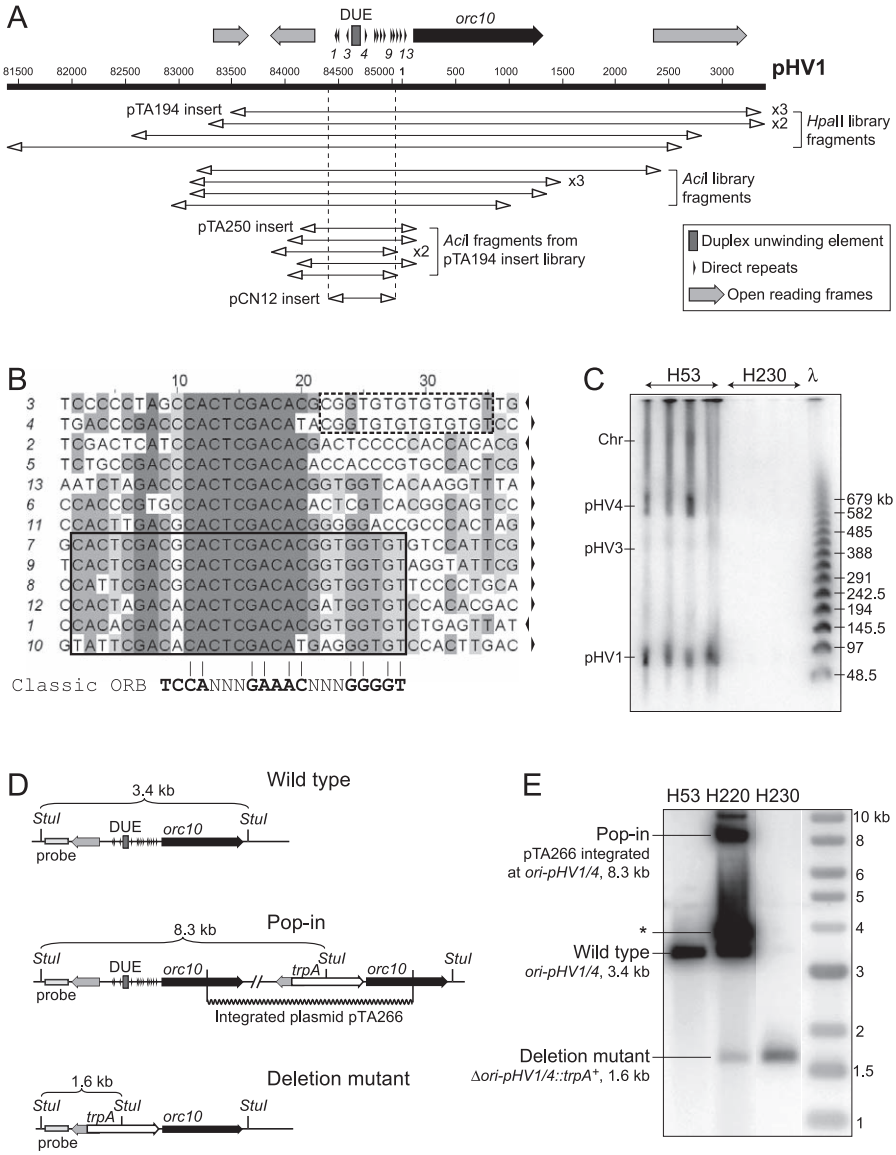


Figure 1. DNA Replication Origin on Contig 454 Is on Chromosomes pHV1 and pHV4

(A) Sequence features of the ARS element isolated from genomic libraries of WR340 DNA. Coordinates of plasmid inserts generated by HpaII and AciI digestion are shown (including insert in pTA194), in addition to the minimal ARS element determined by AciI digestion (in pTA250) or PCR amplification (in pCN12). Numbering refers to TIGR contig 454 (pHV1).

(B) Sequence of repeats found at the intergenic region of the pHV1/4 replication origin (correspond to numbered arrows in Figure 1A). Orientation is indicated by arrows (righthand side) and conserved sequences are shaded. Six of the 13 repeats feature a complete ORB element (boxed), while a core mini-ORB element is conserved in all repeats. The sequence motif found in repeats surrounding the DUE is indicated by the dashed box.

(C) Southern blot of PFG of intact DNA from strains H53 and H230, probed with HpaII ARS insert from pTA194, or intergenic region replaced by *trpA* in pTA266 (see Figure 1D).

(D) Intergenic region of *ori-pHV1/4* was replaced by *trpA* marker by using the deletion construct pTA266. *H. volcanii* H53 was transformed with pTA266 to generate H220, which was used to derive the pHV1/4 origin deletion strain H230. Predicted fragment sizes of StuI digest are indicated.

(E) StuI digest of genomic DNA from strains H53, H220, and H230, probed with DNA flanking the intergenic region. The band indicated * represents episomal DNA carrying the pHV1/4 origin, resulting from excision of the integrated plasmid.

doi:10.1371/journal.pgen.0030077.g001

A total of 35 transformants were obtained. Plasmid DNA from seven transformants was sequenced, and in all cases, the insert corresponded to the same sequence on contig number 454 (Figure 1A). Since *H. volcanii* has more than one replicon, it was striking that only one ARS element was recovered. To ensure this was not a technical artifact of the screen, we repeated the library construction using AciI to generate the partial digest of *H. volcanii* genomic DNA and transformed H49 as before. Plasmid DNA from six AciI library transformants was sequenced, and all contained the same region of the *H. volcanii* genome isolated in the initial screen. This region comprises two divergently transcribed genes including *orc10,* separated by a 910-bp intergenic region featuring a 104-bp A/T-rich putative DUE (68% A/T versus 33% for overall genome) and several direct repeats (Figure 1A). Similar features are encountered at almost all characterized archaeal origins of DNA replication (e.g., [4]). However, it is notable that the nucleotide sequence of the direct repeats

(Figure 1B) bears only little similarity to the ORB identified by Robinson et al. [6].

To determine the minimal sequence needed for DNA replication activity, we carried out a partial AciI digest of a representative 5-kb ARS insert that was cloned in plasmid pTA194 (Figure 1A). Fragments of 1–1.3 kb, 1.3–1.6 kb, 1.6–2 kb, 2–3 kb, and 3–4 kb were excised separately from an agarose gel and ligated with pTA131 to construct an ARS subclone library. Transformants of *H. volcanii* H49 were obtained with DNA from each size range. Plasmid DNA from six transformants in the 1–1.3-kb range was sequenced, and in all cases, the insert included the 910-bp intergenic region (Figure 1A). We were able to delimit the minimal origin further by amplifying a 633-bp fragment of the intergenic region using PCR (Figure 1A). This fragment was subcloned in pTA131 to generate pCN12 and used to transform the *ΔpyrE2 ΔradA* strain H112. Thus, the minimal origin is located in this region, and in contrast to what had been observed for *Halobacterium* sp. NRC-1 [15], the *cdc6/orc1* gene is not required in *cis* for ARS activity in *H. volcanii*.

We used the ARS insert in pTA194 to probe a Southern blot of intact *H. volcanii* DNA displayed on a pulsed field gel (PFG). Two bands were observed (Figure 1C), which correspond in size to replicons pHV4 (690 kb) and pHV1 (86 kb) [14]. The intensity of the two bands was similar, suggesting that this replication origin is present on both pHV1 and pHV4 and that both replicons are present in all cells. We employed the minimal ARS element from the AciI subclone library (in pTA250) to generate low copy–number shuttle vectors with *pyrE2, trpA, hdrB,* and *leuB* selectable markers [21] (Figure S1). The principal advantage of these shuttle vectors over existing plasmids based on the origin of replication from pHV2 is their smaller size (~4.5 kb for pHV1/4-based vectors versus ≥7.5 kb for pHV2-based vectors).

Deletion of Replication Origin from pHV1 and pHV4

The fact that only one ARS element was recovered in the initial genetic screen suggests that this sequence might be "dominant" and thereby prevent the isolation of other origins. We hypothesized that deleting this sequence from pHV1 and pHV4 by a gene knockout system [18,21] would allow the isolation of ARS elements corresponding to origins of DNA replication on the main chromosome and pHV3. The pHV1/4 replication origin and adjacent genes were subcloned to generate pTA252, and a 1-kb fragment containing the intergenic region necessary for ARS activity was replaced by a *trpA* selectable marker for tryptophan biosynthesis [21]. This plasmid (pTA266, Figure 1D) was used to transform the *ΔpyrE2 ΔtrpA* strain H53, and integration at the *orc10* locus was verified by Southern blot (Figure 1E, strain H220). Counter-selection with 5-fluoroorotic acid (5-FOA) was used to ensure loss of integrated pTA266 by intramolecular recombination and to yield a *trpA*-marked deletion of the intergenic region (Figure 1D). While origin deletion events ($trpA^+$ 5-FOA-resistant cells) were obtained, they were outnumbered >1,000-fold by events leading to restoration of the wild-type ($trpA^-$ 5-FOA-resistant cells), indicating a strong bias for maintenance of the replication origin. The deletion was verified by Southern blots of a genomic DNA digest (Figure 1E, strain H230) and intact DNA displayed on a PFG (Figure 1C). The pHV1 and pHV4 replicons are still present (unpublished data) and must therefore be using alternative replication origins. Since the deletion strain H230 did not show any obvious growth defects, these alternative origins presumably act as sites of efficient DNA replication initiation. This observation prompted us to search for other ARS elements.

Isolation of a Chromosomal Origin of DNA Replication

We repeated the screen outlined above using the deletion strain H230 as a source of genomic DNA. AciI was used for a partial digest of H230 genomic DNA, fragments of 3–5 kb were cloned in pTA131, and H49 was transformed as before. Plasmid DNA from nine transformants was sequenced, and in all cases the insert localized to a single region of the *H. volcanii* genome (contig number 455, which corresponds to the main chromosome). This region (Figure 2A) comprises two divergently transcribed genes including *orc1,* separated by a 1,360-bp intergenic region featuring two A/T-rich DUEs and several direct repeats. The nucleotide sequence of the direct repeats (Figure 2B) is 89% identical to the ORB sequence (euryarchaeal consensus 5′-GTTCCAGTGGAAAC-AAA----GGGGG-3′) [6]. In addition to *orc1,* a number of other genes related to DNA replication and repair is found in the vicinity of the ARS (Figure 2A), encoding the DP1 exonuclease subunit of the archaeal D family DNA polymerase [23], the Hef helicase/endonuclease [24], a homologue of the bacterial UvrC nucleotide excision repair protein, an NAD-dependent DNA ligase [25], and the Hel308 helicase [26].

The ARS insert in one plasmid (pTA313) was used to probe a Southern blot of intact *H. volcanii* DN5A on a PFG. As expected, one band was observed (Figure 2C) that corresponds in size to the main chromosome (2.9 Mb) [14]. A 1.1-kb Sau3AI-HindIII fragment of pTA313, comprising the intergenic region only (Figure 2A), was subcloned in pTA131 to generate pTA441 and used to transform H49. Transformants were obtained with high efficiency indicating that, as with the pHV1/4 origin, the *cdc6/orc1* gene is not required in *cis* for ARS activity. To delimit the minimal origin further, a 692-bp fragment of the intergenic region was amplified by PCR (Figure 2A), subcloned in pTA131 to generate pCN11, and used to transform the *ΔpyrE2 ΔradA* strain H112. The sequence was able to maintain the plasmid.

ARS Plasmids with pHV1/4 Replication Origin Are Dominant to, but Less Stable Than, Plasmids with Chromosomal Origin

To explain why only one ARS element was recovered in the initial genetic screen, we determined whether the replication origin located on pHV1 and pHV4 might be "dominant" to other origins. The *ΔpyrE2 ΔradA* strain H112 was transformed with an equimolar mixture of pTA194 and pTA313 (0.5 μg each). These are ARS plasmids from the initial and secondary genetic screens, which carry the pHV1/4 and chromosomal origins, respectively. A total of 23 transformants were analyzed by Southern blotting for the presence of each ARS plasmid (Figure 3). The vast majority (21/23) contained the *ori-pHV1/4* plasmid pTA194 and in most cases (17/23) this was the sole plasmid detectable. Only six transformants contained the *oriC* plasmid pTA313, and in most cases (4/6) pTA194 was also present. To determine the fate of ARS plasmids in transformants containing both pTA194 and pTA313, cells were propagated by restreaking, both while maintaining selection for the *pyrE2* marker and without selection. In all three cases

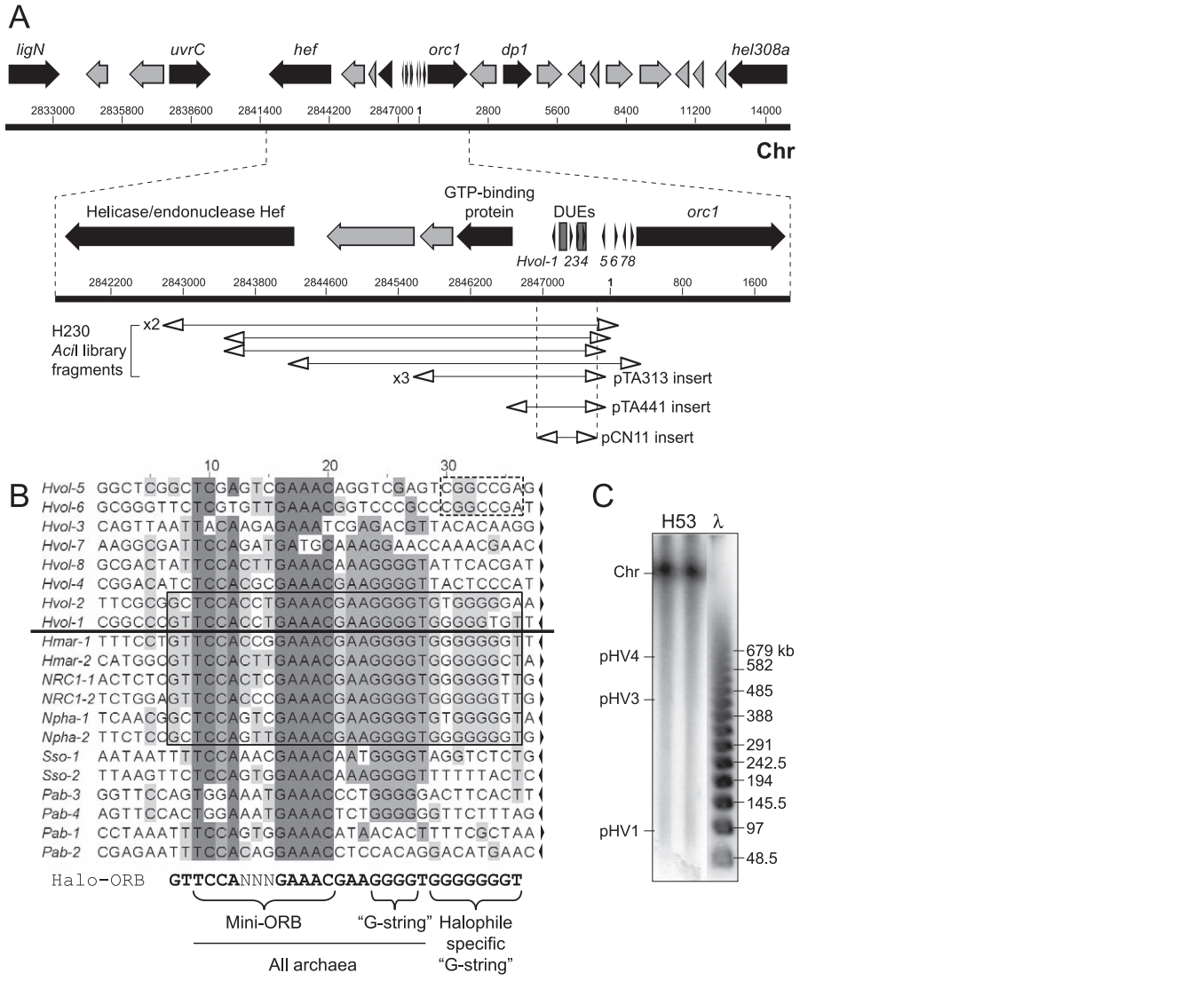


Figure 2. DNA Replication Origin on the Main Chromosome: *oriC1*

(A) Sequence features of the ARS element isolated from genomic libraries of H230 DNA, including selected genes (see text for details). Coordinates of plasmid inserts generated by AciI digestion are shown (including insert in pTA313), in addition to the minimal ARS element in pTA441 and pCN11. See Figure 1A for key. Numbering refers to TIGR contig number 455. Main chromosome, Chr.

(B) Above the line are sequences of repeats found at the intergenic region of the *H. volcanii* chromosomal replication origin (correspond to numbered arrows in Figure 2A). Below the line are sequences of repeats found at other (presumed) archaeal origins. The species and relevant *cdc6/orc1* genes are *H. marismortui cdc6–4 (Hmar-1–2), Halobacterium* sp. NRC-1 *orc7 (NRC1-1-2), Natronomonas pharaonis cdc6–1 (Npha-1–2), S. solfataricus cdc6–1 (Sso-1–2),* and *P. abyssi cdc6 (Pab-1–4).* The orientation is indicated by arrows and conserved positions are shaded. Among halophilic archaea, repeats surrounding the primary DUE feature a longer consensus sequence (Halo-ORB, boxed), which contains the core mini-ORB and "G-string" elements also found in other archaea, plus a halophile-specific "G-string."

(C) Southern blot of PFG of DNA from strain H53, probed with the AciI ARS insert from pTA313.

doi:10.1371/journal.pgen.0030077.g002

analyzed (Figure 3B, transformants 3, 5, 10), only pTA194 remained after further propagation. This was not due to greater stability of pTA194 relative to pTA313, since transformants propagated without selection showed a marked loss of pTA194, whereas pTA313 was well maintained in the transformant where it was present exclusively (Figure 3C, transformant 8). These results recapitulate the outcome of the genetic screens and indicate that the *ori-pHV1/4* plasmid pTA194 is dominant to the *oriC* plasmid pTA313.

The stability of ARS plasmids in the absence of selection was investigated in a quantitative manner. Cultures of H112 transformed with pTA250 or pTA441, which carry the intergenic region of *ori-pHV1/4* and *oriC,* respectively, were propagated for ~25 generations in nonselective Hv-YPC broth. At regular intervals, the fraction of uracil+ cells (indicative of ARS plasmids) was determined by plating on selective and nonselective media. As shown in Figure 3D, the *ori-pHV1/4* plasmid pTA250 is significantly less stable than the *oriC* plasmid pTA441; plasmid loss was calculated to be 16% per generation for pTA250 and 9% per generation for pTA441.

GC-Skew Analyses Predict Additional Replication Origins

The leading and lagging strands of bacterial and archaeal genomes differ in their base composition [27,28], and a

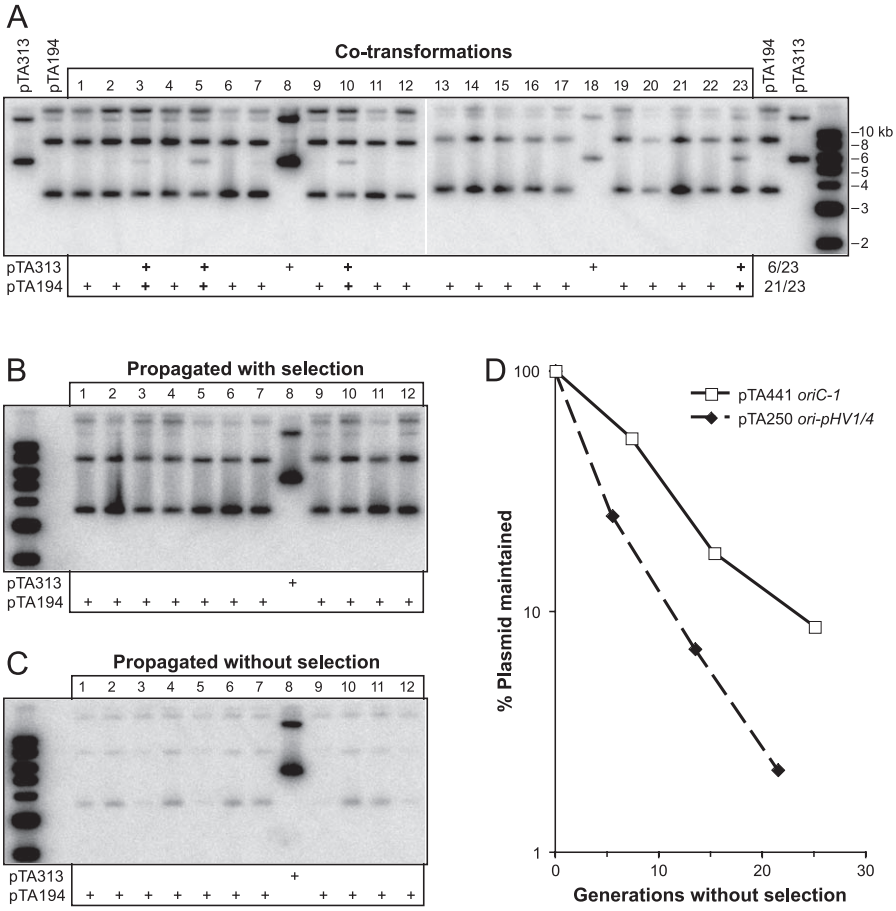


Figure 3. *ori-pHV1/4* ARS Plasmids Are Dominant to, but Less Stable Than, *oriC-1* ARS Plasmids

(A) H112 was transformed with a mixture of 0.5 μg each of pTA194 and pTA313. Crude DNA was isolated from 23 transformants and a Southern blot probed with *pyrE2* sequences.

(B) Transformants 1–12 were propagated by restreaking on selective Hv-Ca agar. DNA was isolated and probed as above.

(C) Transformants 1–12 were restreaked on nonselective Hv-YPC agar. DNA was isolated and probed as above.

(D) H112 containing pTA250 or pTA441 was grown in nonselective Hv-YPC broth. At regular intervals aliquots were plated on selective (Hv-Ca) and nonselective (Hv-YPC) agar, to determine the fraction of uracil+ cells.

doi:10.1371/journal.pgen.0030077.g003

surplus of G over C is usually found on the leading strands of replication. Thus, abrupt changes in the strand-specific nucleotide compositions indicate the presence of a replication origin or terminus. We analyzed the G/C disparity of the largest contig (number 455) of the *H. volcanii* genome sequence, which corresponds in size to the main chromosome (2.9 Mbp). Two different algorithms, ORIGINX [29] (unpublished data) and Z-CURVE (Figure 4A) [30] gave similar results. The "GC-skew" of *H. volcanii* suggests the presence of multiple origins on the main chromosome, similar to what has been proposed for *Halobacterium* sp. NRC-1 [30,31]. The amplitude of the calculated *H. volcanii* GC-skew was similar to that of *H. marismortui,* but substantially smaller than was observed for the normalized GC-skews of the main chromosomes of three other halophilic archaea and a non-halophilic archaeon (Figure 4B). Whereas the *oriC* of the halophilic bacterium *Salinibacter ruber* has a well-defined GC-skew minimum at the origin, the four halophilic archaea show a GC-skew with inverted polarity, when compared to bacteria and thermophilic archaea. Therefore, the leading strand of replication contains an excess of C in haloarchaea, similar to what has been suggested for *Mycoplasma* species [32]. While the mechanisms that establish these skews at a molecular level are still poorly understood, it is of note that their formation seems to be independent of GC% of genomes [32]. For example, the chromosome of the haloarchaeon *Haloquadratum walsbyi,* which is not GC-rich (48% GC), also shows an inverted polarity.

The *H. volcanii* G/C disparity curve has three peaks for the main chromosome (Figure 4A): (1) A peak near the ARS element identified in genomic libraries from H230, which is associated with the *orc1* gene *(oriC1);* (2) A peak associated with the *orc11* and *orc14* genes. This chromosomal region has a base composition different than the rest of the genome and contains putative viral genes such as HNHc endonuclease, VirB4, and tragVirD4 helicases or bacteriophage T4-like integrase (unpublished data). Moreover, it does not coincide with any ARS elements isolated in this work; and (3) A peak associated with the *orc5* gene, suggesting the presence of a second replication origin on the main chromosome *(oriC2).* GC, AT, and G + C content disparities were also carried out for contigs corresponding to the smaller replicons and predicted the location of DNA replication origins for pHV1 (confirming the origin identified in the first ARS screen),

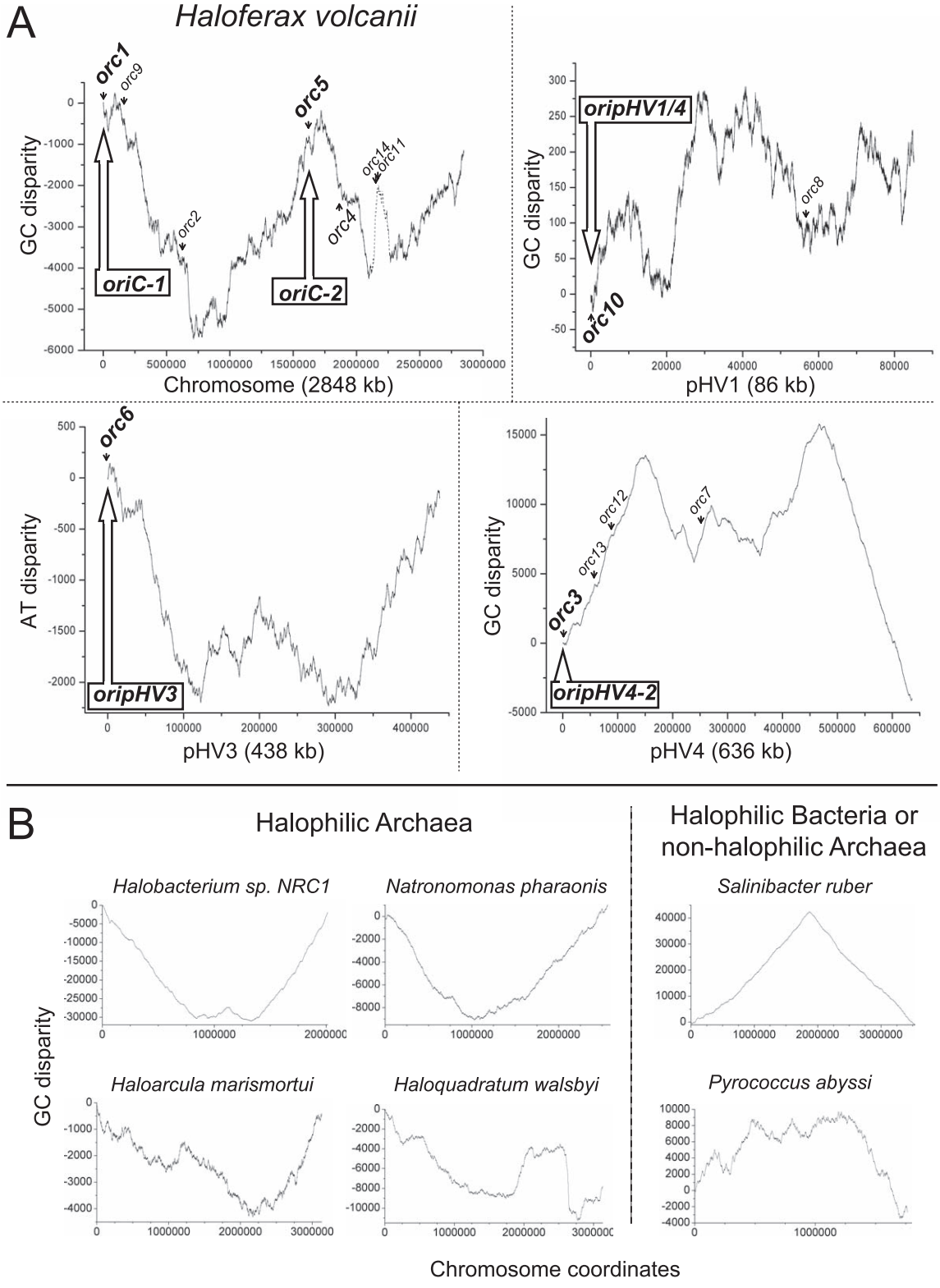


Figure 4. Nucleotide Disparity Curves

(A) Nucleotide disparity curves of *H. volcanii* DS2 genome sequence. Positions of the 14 *cdc6*/*orc1* genes and five replication origins identified in this work are indicated. A putative prophage sequence with high A + T content results in a sharp peak of nucleotide disparity for the main chromosome (indicated by dotted line), but does not coincide with any ARS element isolated here. Nucleotide skews are calculated using the ZPLOTTER program (http://tubic.tju.edu.cn/zcurve).

(B) GC disparity curves of the chromosome of four halophilic archaea, a halophilic bacterium, and a non-halophilic archaeon. Chromosome coordinates have been offset to begin at *H.* sp. NRC-1 *orc7, N. pharaonis cdc6, H. marismortui cdc6–4, H. walsbyi cdc6–1, S. ruber dnaA,* and *P. abyssi cdc6* genes, which are assumed or proven to be associated with the replication origins. The halophilic bacterium and non-halophilic archaeon were used as a control.

doi:10.1371/journal.pgen.0030077.g004

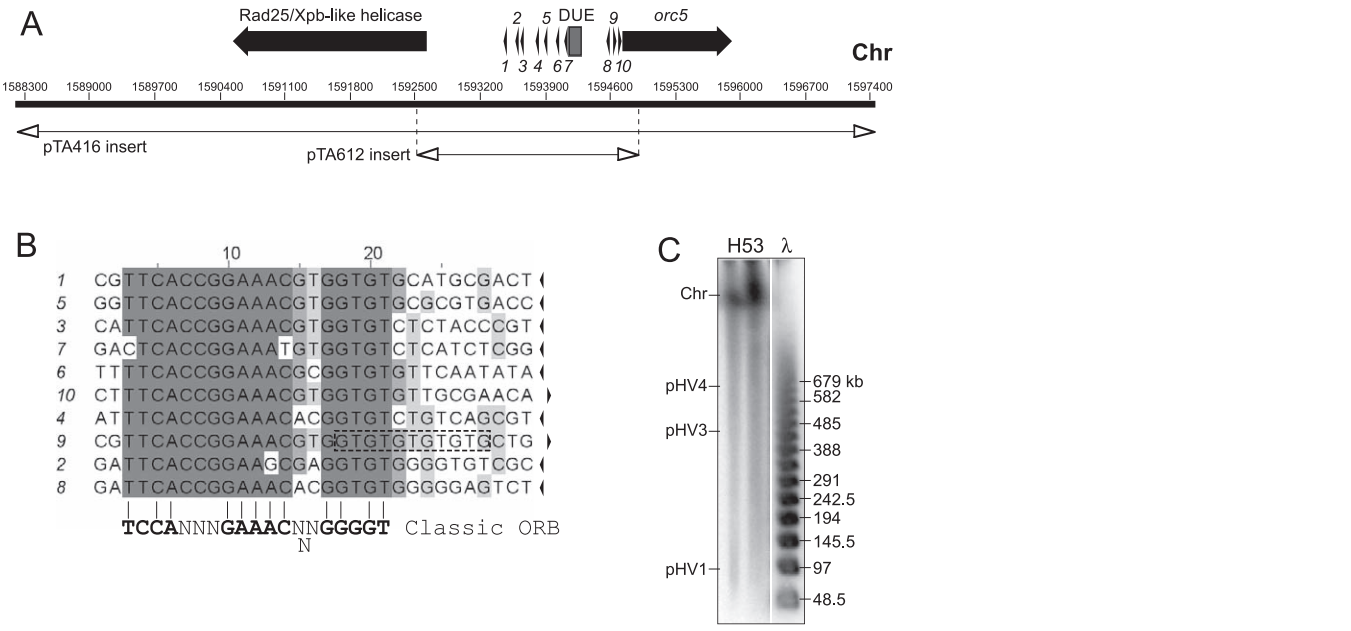


Figure 5. A Second DNA Replication Origin on the Main Chromosome: *oriC2*

(A) Sequence features of the second origin from main chromosome. Coordinates of the 9.3-kb NotI genomic fragment cloned in pTA416 and the 2.4-kb intergenic region cloned in pTA612 are shown. See Figure 1A for key. Numbering refers to TIGR contig number 455. Main chromosome, Chr.

(B) Sequence of repeats found at the intergenic region of *oriC2* (numbered arrows in Figure 5A). The orientation is indicated by arrows and conserved sequences are shaded.

(C) Southern blot of PFG of DNA from strain H53, probed with intergenic region cloned in pTA612.

doi:10.1371/journal.pgen.0030077.g005

pHV3, and a second origin on pHV4 (Figure 4A); all of these peaks are located near *cdc6/orc1* genes. Strikingly, *ori-pHV1/4* sequence could only be found on contig number 454 (pHV1) and not on contig number 452 (which should correspond in size to pHV4).

To test the possibility of a second chromosomal origin and to identify origins on the smaller replicons, we directly cloned candidate ARS elements. A region carrying the putative second chromosomal replication origin *(oriC2)* was isolated as a 9.3-kb NotI fragment from a genomic DNA library and was cloned in pTA416 (Figure 5A). This fragment contains two divergently transcribed genes coding for Orc5 and a putative Rad25/Xpb-related helicase, separated by a 2,094-bp intergenic region featuring an A/T-rich DUE and several direct repeats with similarity to the ORB consensus (Figure 5B). A 2.4-kb fragment comprising this intergenic region was amplified by PCR, cloned in pTA131 to generate pTA612, and used to transform the *ΔpyrE2 ΔradA* strain H112. Transformants were obtained with high frequency, indicating that this region supports ARS activity (Figure S2). However, after 15 d of incubation the colonies were significantly smaller than those seen with H112 transformed with pTA250 (pHV1/4 origin) or pTA441 *(oriC1)*. This suggests that the efficiency of this putative second chromosomal origin (or the stability of the plasmid) is significantly lower than that of the pHV1/4 and *orc1*-associated origins. Moreover, the intergenic sequence upstream of *orc5,* which contains *oriC2,* can be deleted efficiently by using the same gene knockout procedure as was used to delete the pHV1/4 origin [21]. The resulting strain CN28 shows no growth defect (unpublished data).

The 793-bp intergenic sequence on contig number 453 (corresponding to pHV3) is located between genes coding for a putative TATA box-binding protein and Orc6, and features two DUEs and several direct repeats with similarity to ORB consensus sequence (Figure 6A and 6B). To test if this intergenic region confers ARS activity, a 693-bp PCR fragment including the DUEs and repeats was inserted in pTA131 to generate pCN26 and used to transform H112. Transformants were obtained with high frequency (Figure S2). When probed with the *ori-pHV3* sequence, a PFG confirmed that this origin is located on the pHV3 replicon (Figure 6C).

The 688-bp intergenic region on contig number 452 (which should correspond to pHV4) is located between a putative translation initiation factor and *orc3* genes and features a DUE and multiple direct repeats that are different from the ORB consensus sequence (Figure 6D and 6E). This putative origin was tested for ARS activity by cloning a 592-bp PCR fragment of the intergenic sequence in pTA131 to generate pCN27 and used to transform H112. Significantly fewer transformants were obtained with pCN27 *(ori-pHV4–2)* than with pCN26 *(ori-pHV3)* or pTA250 *(ori-pHV1/4),* indicating that the second origin of pHV4 is not used efficiently (Figure S2). Moreover, a PFG probed with *ori-pHV4–2* sequence showed that this origin is actually located on the 2.9-Mb main chromosome (Figure 6F), suggesting that the strains used in our experiments and/or sequenced by TIGR may have undergone genome rearrangements.

The stability of ARS plasmids pTA612 *(oriC2),* pCN26 *(ori-pHV3),* and pCN27 *(ori-pHV4–2)* was determined in a qualitative manner. Six transformants of H112 containing each ARS plasmid were grown on nonselective agar and then restreaked on selective medium. After restreaking, none of the transformants showed growth on selective agar, indicat-

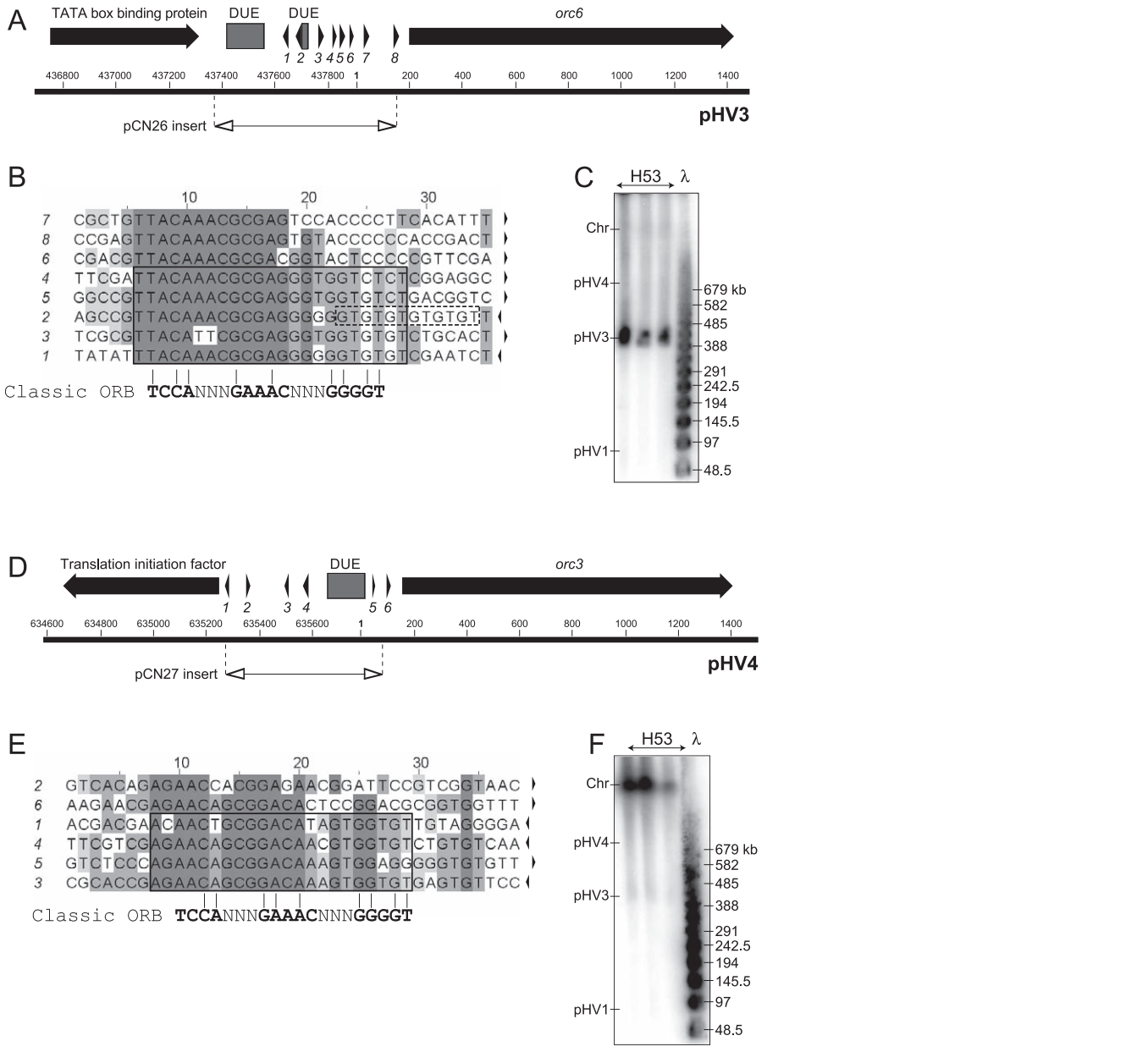


Figure 6. Additional ARS Elements Predicted by Nucleotide Composition

(A) Sequence features of replication origin on pHV3. Coordinates of 693-bp intergenic region cloned in pCN26 are shown. See Figure 1A for key. Numbering refers to TIGR contig number 453 (pHV3).

(B) Sequence of repeats found at the intergenic region of the pHV3 origin (numbered arrows in Figure 6A). The orientation is indicated by arrows and conserved sequences are shaded.

(C) Southern blot of PFG of DNA from strain H53, probed with intergenic region cloned in pCN26.

(D) Sequence features of ARS element in contig number 452. Coordinates of 592-bp intergenic region cloned in pCN27 are shown. Numbering refers to TIGR contig number 452 (pHV4).

(E) Sequence of repeats found at intergenic region of contig number 452 ARS element (numbered arrows in Figure 6D).

(F) Southern blot of PFG of DNA from strain H53, probed with intergenic region cloned in pCN27.

doi:10.1371/journal.pgen.0030077.g006

ing complete loss of the plasmid (Table 2). In comparison, transformants with pTA441 *(oriC1)* showed normal growth on selective agar, indicating efficient maintenance of the plasmid. Transformants with pTA250 *(ori-pHV1/4)* showed growth on selective agar, but the number of colonies was significantly reduced relative to growth on nonselective agar, indicating some loss of the plasmid.

As a control, we tested whether the intergenic region adjacent to *orc4* could function as an ARS element; this region does not coincide with a peak of G/C disparity (Figure 4A) and does not feature a DUE. A 1.7-kb ApaI-BglII genomic DNA fragment of this intergenic region was cloned in pTA131 to generate pTA611 and used to transform H112. No transformants were observed, indicating that not every *cdc6/orc1* gene is associated with a potential DNA replication origin.

Table 2. Characteristics of Replication Origins

Origin	Location		Transformant Colony Size[a]	Transformation Efficiency[b]	Plasmid Stability
	Predicted	Observed			
oriC1	Chr	Chr	Big	$\sim 10^6$	Maintained
oriC2	Chr	Chr	Small	$\sim 10^6$	Lost
ori-pHV1/4	pHV1	pHV1/pHV4	Big	$\sim 10^6$	Partial
ori-pHV3	pHV3	pHV3	Small	$\sim 10^6$	Lost
ori-pHV4–2	pHV4	Chr	Small	~100	Lost

[a]H112 transformed with ARS plasmid, Figure S2.
[b]Colony-forming units/μg DNA.
doi:10.1371/journal.pgen.0030077.t002

ARS Elements Function as Replication Origins

RIP mapping [33] was used to determine the exact positions of DNA replication initiation at the origins we had identified. A preparation of enriched nascent DNA strands was used in primer extension reactions using ^{32}P-labeled primers hybridizing to the leading strand either side of the DUEs of each ARS element (Table S1). The sizes of the amplification products were determined using electrophoresis under denaturing conditions (Figure 7A), allowing identification of the shortest amplification products that mark the start site of the leading strand. Clear transition points between leading and lagging strand replication within or near the DUEs were revealed for both strands of each origin (Figure 7B). It is of note that the detected transition points colocalize with chromosomal regions where DNA is predicted to be bent. Amplification products obtained in negative control experiments, which used linearized plasmids containing the origin sequences that were isolated from *Escherichia coli,* are indicated in Figure 7B using dots. As expected, they do not coincide with the extension products obtained using *H. volcanii* replication intermediates. These results indicate that all the ARS elements isolated during this work are functional in their chromosomal context.

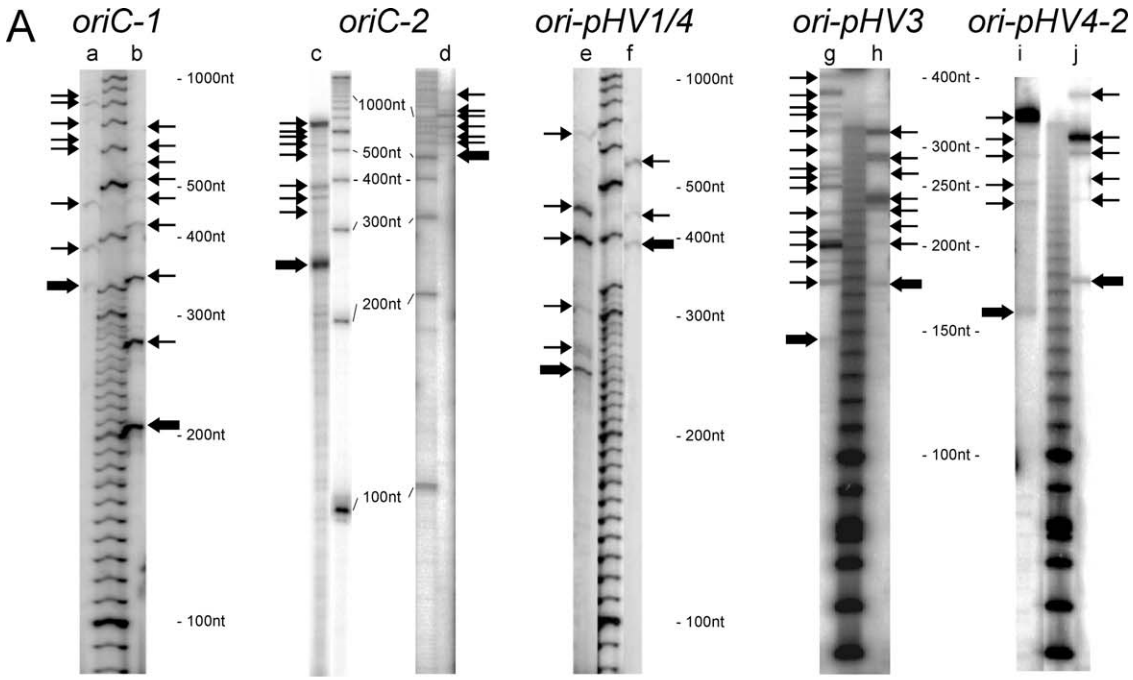
Discussion

Initiation of DNA replication at an origin depends on *cis*-acting sequence elements. In bacteria and unicellular eukaryotes, these have been well-characterized using genetics. In contrast, archaeal replication origins were identified relatively recently [2], and only limited genetic methods have been used to study DNA replication in archaea [15]. In this work, we isolated several ARS elements that use chromosomally encoded factors for their replication. RIP assays were used to confirm that these ARS elements correspond to functional replication origins in their chromosomal context. The origins are distributed on the different replicons of *H. volcanii,* including two on the main chromosome. Previous work in *Sulfolobus solfataricus* and recent studies in *A. pernix* have shown that Crenarchaeota can use more than one origin to replicate a circular chromosome [5,6]. Bioinformatics has suggested that *Halobacterium* sp. NRC-1, a euryarchaeon like *H. volcanii,* might have two chromosomal replication origins [34], but attempts to identify the second origin experimentally were not successful [15]. Our work provides the first proven example from the Euryarchaeota of multiple origins per replicon.

The general characteristics of the *H. volcanii* replication origins are similar. All contain AT-rich sequences required for unwinding and/or bending of the origin DNA. These are surrounded by repeated sequence motifs that correspond to the classic ORB elements found at other archaeal origins. However, some of these repeats show significant differences to the archaeal mini-ORB consensus. Many archaeal ORBs also feature a characteristic "G-string" element (Figure 2) that contributes to Cdc6/Orc1 binding at the origin (S. Bell, personal communication). While "G-string" elements are found at all origins of *H. volcanii,* the two ORBs surrounding the primary DUE of *oriC1* also contain an extended "G-string" that appears to be specific to halophiles (Figure 2). Furthermore, long GT-stretches are found after some repeats, such as the two ORBs surrounding the DUE of *ori-pHV1/4* (Figure 1B). The subtle variations found in halophilic ORB sequences might contribute to origin-specific binding by different Cdc6/Orc1 proteins. Alternatively, these halophile-specific features could promote DNA bending at high intracellular salt concentrations, thus favoring the formation of a higher order complex between the origin and initiator proteins [9].

The number of Cdc6/Orc1 proteins is almost certainly higher than the number of origins in *H. volcanii;* for example, the intergenic region adjacent to *orc4* does not show ARS activity. Furthermore, we were able to delete several *cdc6/orc1* genes (*orc1, orc5,* and *orc10,* unpublished data), suggesting that their functions at least partially overlap. It is possible that some Cdc6/Orc1 proteins might promote the initiation of "routine" DNA replication, while other initiators are adapted to function under specific physiological conditions. For example, our observation that the second chromosomal origin *(oriC2)* appears to function less efficiently in laboratory conditions suggests that it is only used under certain circumstances. This could be due to location of the origins on ARS plasmids, as opposed to their native chromosomal loci. On the other hand, it is noteworthy that the overall pI values of the different Cdc6/Orc1 proteins vary between 3.98 and 5.59 (Table 1), with the *oriC1*-associated Orc1 being the most acidic, raising the possibility that they might function optimally under different salt concentrations.

A surprising result that emerged from our initial genetic screen for ARS elements was that only one origin *(ori-pHV1/4)* was isolated. This is unlikely to be due to technical limitations of the method, since extensive libraries were constructed using different restriction enzymes. Moreover, we were able


A
oriC-1
oriC-2
ori-pHV1/4
ori-pHV3
ori-pHV4-2
a b
c d
e f
g h
i j
- 1000nt
- 500nt
- 400nt
- 300nt
- 200nt
- 100nt
1000nt
500nt
400nt
300nt
200nt
100nt
- 400nt -
- 300nt -
- 250nt -
- 200nt -
- 150nt -
- 100nt -

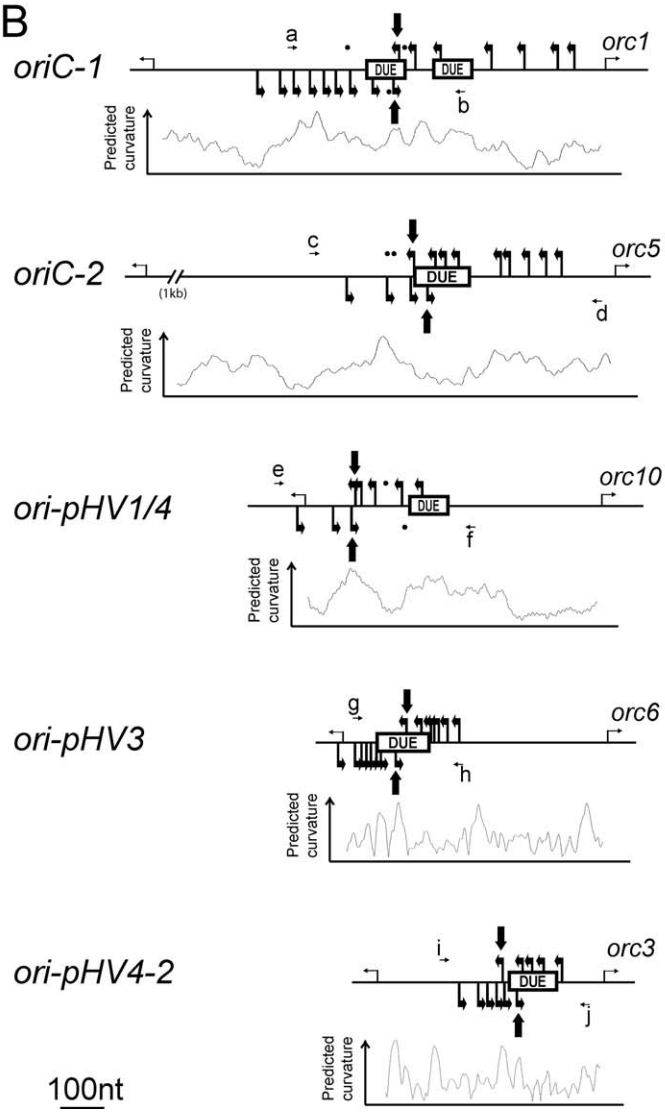

B
oriC-1
oriC-2
ori-pHV1/4
ori-pHV3
ori-pHV4-2
orc1
orc5
orc10
orc6
orc3
DUE
(1kb)
Predicted curvature
100nt

Figure 7. Mapping of Initiation Sites at the Replication Origins

(A) Primer extension reactions were performed with enriched replicating intermediate DNA fragments and separated on a denaturing polyacrylamide gel. Corresponding primers are indicated on top of the gels (see Table S1 for details), arrows refer to the 5′ end of amplification products and transition points are indicated by bold arrows.

(B) Primers and detected initiation sites along the sequence. The positions of DUEs, the associated *cdc6/orc1* gene on the right, and the upstream open reading frame on the left are shown for each replication origin. Dots indicate weak amplification products obtained with control templates (linearized plasmid DNA carrying the replication origins, isolated from *E. coli* cells).

doi:10.1371/journal.pgen.0030077.g007

to recapitulate the outcome of the screen by cotransforming *H. volcanii* with a mixture of *ori-pHV1/4* and *oriC1* ARS plasmids (Figure 3). The majority (74%) of transformants contained only the *ori-pHV1/4* plasmid, and less than 9% contained only the *oriC1* plasmid (Figure 3A). This disparity is unlikely to be due to differences in plasmid establishment, since the two ARS constructs show the same transformation efficiency (Table 2) and confer a similar phenotypic load on their host (Figure S2). Instead, it would appear that the *ori-pHV1/4* ARS plasmid is dominant to the *oriC1* ARS plasmid. Evidence in favor of this suggestion emerged from the examination of transformants containing both plasmids (Figure 3B). Upon further propagation, only the *ori-pHV1/4* ARS plasmid remained while the *oriC1* ARS plasmid was lost. This result is all the more intriguing given that *ori-pHV1/4* plasmids are significantly less stable than *oriC1* plasmids (Figure 3C and 3D). Furthermore, the copy number of *ori-pHV1/4* plasmids is lower than that of *oriC1* plasmids (unpublished data). Thus, our results cannot be explained by mere incompatibility of ARS plasmids, since this would favor the retention of the *oriC1* plasmid.

Instead, we suggest that in *H. volcanii* there is a functional hierarchy of replication origins, which arises from competition for common replication factors. "Dominant" origins such as *ori-pHV1/4* might be bound more efficiently by Cdc6/Orc1 or other proteins such as MCM helicase and primase that participate in replication initiation. Other factors might also contribute to differential usage of origins. ARS elements that initiate efficiently can be selectively enriched from a genomic library of *Saccharomyces cerevisiae* [35], and reduced initiation efficiency was found to correlate with transcriptional activity directed towards the ARS element. This suggests that transcriptional interference with the prereplication complex is a determinant of origin hierarchy in eukaryotes. It is noteworthy that four of the five origins identified in this work (and almost all other archaeal origins) feature transcription units that are orientated away from replication initiation sites (Figures 1, 2, 5, and 6). Therefore, transcriptional interference might also regulate origin firing in archaea.

Our genetic observations indicate that while the functions of initiator proteins at least partially overlap in *H. volcanii,* the replication origins *oriC1* and *ori-pHV1/4* play a key role in the replication of their corresponding chromosomes. This suggests that duplication of replication origins, and not of the initiator genes, could have allowed the *H. volcanii* genome to develop toward a multireplicon structure. This process might still be continuing, since there are important differences between our data and the assembled genome sequence. For instance, both pHV1 and pHV4 carry a similar (if not identical) replication origin. A comparable situation is found with *Halobacterium* sp. NRC-1 plasmids pNRC100 and pNRC200, which share a large region of around 150 kb including long (33 and 39 kb) inverted repeats [13]. An additional example is *ori-pHV4–2,* which in the current version of the assembled genome sequence is located on the contig corresponding to pHV4, while in our laboratory strains is carried on the main chromosome. Detailed manual examination of the genome sequencing data (including shotgun data and closure results) indicate that the assembly is an accurate representation of the strain used for sequencing. Though it is possible the assembly is inaccurate, we believe it is more likely that there are genomic differences between the strain sequenced and strains used in our experiments. For instance, the *H. volcanii* DS2 strain was used to isolate DNA for sequencing (Table 3), whereas genetic and biochemical experiments used derivatives of *H. volcanii* WFD11 or DS70 [19,20].

These discrepancies suggest that *H. volcanii,* whose genome is considered relatively stable compared to other haloarchaea, undergoes periodic genome arrangements that are possibly mediated by recombination at the replication origins. In this respect, it is noteworthy that the two rRNA operons are located close to the two chromosomal origins described in this work. The *rrnA* operon is 200 kb away from *oriC1,* and *rrnB* is only 6 kb away from *oriC2.* Homologous recombination between identical sequences at these *rrn* operons could provide a mechanism for origin movement throughout the genome. Intriguingly, the *rrn* operons are both oriented away from the origins, presumably to avoid collisions between replication fork and rRNA transcription machinery.

It is notable that the *H. volcanii* GC-skew plots do not delineate origins as clearly as the plots of the other sequenced haloarchaea (Figure 4). For example, the peak in contig 454 does not match the ARS isolated in the initial screen. Conversely, there is a peak in contig 455 that has higher amplitude than the two origins detected but is not associated with an experimentally determined origin (Figure 4A); this

Table 3. *H. volcanii* Strains

Strain	Genotype	Background	Derivation[a]
DS2	Wild-type	—	[40]
WR340	*his*	WFD11	[18]
WR480	*his ΔpyrE2*	WFD11	[18]
H26	*ΔpyrE2*	DS70	[21]
H49	*his ΔpyrE2 ΔradA*	WFD11	WR480 *ΔradA*
H53	*ΔpyrE2 ΔtrpA*	DS70	[21]
H112	*ΔpyrE2 ΔradA*	DS70	H26 *ΔradA*
H220	*ΔpyrE2 ΔtrpA ori-pHV1/4*::pTA266	DS70	H53, pTA266 integrated
H230	*ΔpyrE2 ΔtrpA Δori-pHV1/4::trpA*$^+$	DS70	H220, pTA266 excised

[a]Unless indicated otherwise, source of strains was this study. The DS2 strain used for genome sequencing was obtained directly from ATCC.

doi:10.1371/journal.pgen.0030077.t003

Table 4. Plasmids

Plasmid	Relevant Properties
pTA131	Integrative vector based on pBluescript II, with *pyrE2* marker [21]
pTA194	pTA131 containing 5-kb HpaII genomic fragment from WR340 with pHV1/4 replication origin inserted at ClaI site
pTA250	pTA131 containing 1.1-kb AciI fragment of pTA194 insert with intergenic region of pHV1/4 replication origin inserted at ClaI site
pTA252	pTA131 containing 3-kb BamHI-StuI fragment of pTA194 with pHV1/4 replication origin inserted at BamHI and NotI sites
pTA266	pTA252 with deletion of 1-kb BsmI-XbaI fragment containing intergenic region of pHV1/4 replication origin, replaced by 1-kb BamHI-XbaI fragment of pTA106 containing *trpA* marker [21]
pTA313	pTA131 containing 2.2-kb AciI genomic fragment from H230 with chromosomal replication origin *oriC1* inserted at ClaI site
pTA351	*E. coli/H. volcanii* shuttle vector with *trpA* marker derived from pTA132 [21]. Contains 948-bp BmgBI-EcoRV fragment of pTA250 with pHV1/4 replication origin inserted at PciI site
pTA352	*E. coli/H. volcanii* shuttle vector with *leuB* marker derived from pTA133 [21]. Contains 948-bp BmgBI-EcoRV fragment of pTA250 with pHV1/4 replication origin inserted at PciI site
pTA354	*E. coli/H. volcanii* shuttle vector with *pyrE2* marker derived from pTA131 [21]. Contains 948-bp BmgBI-EcoRV fragment of pTA250 with pHV1/4 replication origin inserted at PciI site
pTA356	*E. coli/H. volcanii* shuttle vector with *hdrB* marker derived from pTA192 [21]. Contains 948-bp BmgBI-EcoRV fragment of pTA250 with pHV1/4 replication origin inserted at PciI site
pTA416	pBluescript II containing 9.3-kb NotI genomic DNA fragment from WR340 with chromosomal origin *oriC2* inserted at NotI site
pTA441	pTA131 containing 1.1-kb Sau3AI-HindIII fragment of pTA313 with intergenic region of *oriC1* inserted at BamHI and HindIII sites
pTA611	pTA131 containing 1.7-kb fragment ApaI-BglII genomic DNA fragment from WR340 with intergenic region adjacent to *orc4* inserted at ApaI and BamHI sites
pTA612	pTA131 containing 2.4-kb fragment of pTA416 with intergenic region of *oriC2,* amplified by PCR using F-1cdc and R-1radh primers and inserted at EcoRV site
pCN11	pTA131 containing 692-bp fragment of the intergenic region of *oriC1,* amplified by PCR using OCN33 and OCN35 primers and inserted at NotI site
pCN12	pTA131 containing 633-bp fragment of pTA194 with intergenic region of pHV1/4 replication origin, amplified by PCR using OCN27 and OCN28 primers and inserted at NotI site
pCN26	pTA131 containing 693-bp fragment of the intergenic region of pHV3 replication origin, amplified by PCR using OCN180 and OCN181 primers and inserted at BamHI and EcoRI sites
pCN27	pTA131 containing 593-bp fragment of the intergenic region of contig 452 replication origin, amplified by PCR using OCN182 and OCN183 primers and inserted at BamHI and EcoRI sites

doi:10.1371/journal.pgen.0030077.t004

peak is most likely due to a recent integration by an AT-rich prophage. These quirks could indicate either that the *H. volcanii* "origin signal" (i.e., the skew caused by origin usage) is weak, or that the origin is mobile and has left behind a residual signal at its previous location. Finally, the fact that haloarchaeal genomes present "inverted" GC-skews indicates that the haloarchaea may not follow the same rules for genome structure patterning as in other species.

Given the apparent plasticity of the *H. volcanii* genome, we believe that pHV4, pHV3, and pHV1 should be considered as secondary chromosomes and not plasmids, in the sense that a plasmid is an extrachromosomal element and is often characterized by "selfish" behavior. Prokaryotic chromosome biology has been dominated by the archetype of *E. coli,* but there are many examples of bacteria with more than one chromosome [36]; like *H. volcanii,* they often show evidence of dynamic rearrangements between the replicons (e.g., [37]). The distinction between mini-chromosomes and mega-plasmids is usually made on the basis of replicon size and the presence of essential (housekeeping) genes. In this regard, both pHV4 and pHV3 are large (690 and 442 kb, respectively), and while genes predicted to be essential are only found on the former, *H. volcanii* strains that have lost pHV3 grow very slowly and filament [20]. More relevant to this work, pHV4, pHV3, and pHV1 all use Cdc6/Orc1-dependent origins for replication, as does the main chromosome. By contrast, pHV2 is very small (6.4 kb), lacks homology with chromosomal sequences, is easily cured, and uses a distinct (presumably Rep-dependent) replication origin [19]. Thus it is a plasmid.

In conclusion, we have shown that the four chromosomes of *H. volcanii* are replicated in an analogous manner using Cdc6/Orc1-dependent origins. Since efficient tools now exist to characterize the *cis*-acting requirements for replication initiation in archaea, *H. volcanii* will serve as an interesting model system for studies on the regulated replication of complex archaeal genomes.

Materials and Methods

Unless stated otherwise, chemicals were from Sigma (http://www.sigma-aldrich.com) and enzymes from New England Biolabs (http://www.neb.com). Standard molecular techniques were used [38].

Strains and plasmids. *H. volcanii* strains are shown in Table 3, plasmids in Table 4, and oligonucleotides in Table S1.

Media and growth conditions. *H. volcanii* strains were grown at 45 °C on either complete (Hv-YPC) or casamino acids (Hv-Ca) agar, or in Hv-YPC or casamino acids broth, as described previously [21,39].

Molecular genetic methods. Transformation of *H. volcanii* and isolation of total genomic DNA were carried out as described previously [21]. To isolate plasmid DNA, 2 ml of a saturated culture (grown in Hv-Ca broth) was centrifuged at 3,300 ×*g* for 8 min and the cells were resuspended in 50 μl of 1 M NaCl, 20 mM Tris HCl (pH 7.5); thereafter, a QIAprep miniprep kit (Qiagen, http://www.qiagen.com) was used according to the manufacturer's instructions. To isolate crude DNA, cells were resuspended in 400 μl of water and incubated at 70 °C for 10 min.

Genomic DNA libraries. *H. volcanii* DNA was digested with HpaII, AciI, or TaqI for 30 min at the recommended temperature, using ~0.2 units of enzyme/μg DNA in suboptimal buffer (e.g., New England Biolabs buffer 1 for AciI). DNA fragments of the desired size were excised from agarose gels and ligated with plasmid pTA131 [21], which had previously been cut to completion with ClaI and the DNA ends dephosphorylated. The plasmid library was used to transform an *E. coli dam*$^-$ strain, and DNA was prepared directly from colonies to avoid differential amplification. Coverage of the libraries was >99.5% (>5,000 *E. coli* colonies were used in DNA preparation and *H. volcanii* has a genome size of ~4 Mb). DNA was used to transform *H. volcanii pyrE2 radA* mutants H49 or H112, and transformants were selected on Hv-Ca plates lacking uracil.

Deletion of pHV1/4 origin. A 3-kb BamHI-StuI fragment of pTA194 with the pHV1/4 replication origin was subcloned to generate pTA252, and a 1-kb BsmI-XbaI fragment of the intergenic region was replaced by a *trpA* selectable marker [21] to generate pTA266 (Figure 2D). pTA266 was used to transform *ΔpyrE2 ΔtrpA* strain H53, and transformants were selected on Hv-Ca plates lacking uracil and tryptophan. One transformant (H220) was grown without selection for ~30 generations and plated on Hv-Ca + 5-FOA (5-fluoroorotic acid) to select for loss of the integrated plasmid. Around 0.4% of cells in the culture were 5-FOA-resistant but only 0.13% of these were Trp$^+$ and therefore had deleted the pHV1/4 replication origin.

Estimation of ARS plasmid stability. H112 containing pTA250 or pTA441 (maintained by selection on Hv-Ca agar) was used to inoculate a culture in Hv-YPC broth and grown at 45 °C. At regular intervals aliquots were plated Hv-YPC agar, and colonies patched on Hv-Ca to determine the fraction of uracil+ cells. % plasmid loss per generation (*l*) was calculated using the formula $l = (1 - \sqrt[g]{ura^+}) \times 100$

where n is the number of generations and ura^+ is the fraction of uracil+ cells.

PFG electrophoresis. Intact *H. volcanii* DNA was prepared in agarose plugs. 2 ml of culture (OD_{650nm} of 1.0) was pelleted at 3,300 ×*g* for 10 min, 4 °C, resuspended in 1 ml of cold spheroplasting solution (15% sucrose, 1 M NaCl, 27 mM KCl, 50 mM Tris-HCl [pH 8.5]) + 0.1% NaN_3, and pelleted again. Cells were gently resuspended in 80 µl of spheroplasting solution, transferred to 42 °C, mixed with 100 µl of 1.5% low-melt agarose (in 0.5× spheroplasting solution, 100 mM EDTA) and pipetted into plug moulds (Bio-Rad, http://www.bio-rad.com). Plugs were incubated in 5 ml of lysis solution (1% sarkosyl, 500 mM EDTA, 20 mM Tris-HCl [pH 8.8]) + proteinase K (0.5 mg/ml) for 3 h at 52 °C, then transferred to fresh lysis buffer + proteinase K + RNaseA (30 mg/ml) and incubated overnight at 52 °C. Plugs were washed three times in 10 ml of 100 mM EDTA, 25 mM Tris-HCl (pH 7.5) at 37 °C, equilibrated in 0.5× Tris-borate-EDTA (TBE) for 90 min at 20 °C and exposed to 50 Gy of γ radiation (^{137}Cs, 375 Gy/s) to linearize circular DNA molecules. Plugs were loaded onto a 1% agarose 0.5× TBE gel and electrophoresis was performed at 14 °C in a CHEF mapper (Bio-Rad) using 0.5× TBE buffer, voltage gradient of 6 V/cm, switch angle of 120°, and switch times of 0.47 s (initial) to 1 min 33.83 s (final). Total run time was 20 h 18 min.

RIP mapping assay. *H. volcanii* DS2 was diluted to an OD_{650nm} of 0.15 in 100 ml of Hv-YPC media, grown to $OD_{650\ nm}$ of 0.3, and pelleted. Cells were resuspended in 4 ml of lysis buffer (25 mM Tris-HCl [pH 7.5], 20 mM EDTA, 100 mM NaCl, 200 µg ml^{-1} proteinase K, 1% SDS). After 1 h incubation at 50 °C, 4 g of CsCl and 100 µl of Hoechst-33342 (5 mg ml^{-1}) were added and the refractive index adjusted to 1.410 with CsCl (1 g ml^{-1}). DNA was purified by CsCl gradient ultracentrifugation. To enrich for replicating intermediates, total DNA was passed down a BND-cellulose column pre-equilibrated with NET buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, and 1 M NaCl) to selectively bind single-stranded DNA. After washing with NET buffer, bound DNA was eluted with NET buffer + 1.8% caffeine at 50 °C. DNA was isopropanol-precipitated and resuspended in TE buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA) at 1 µg $µl^{-1}$. After phosphorylation of 5′-OH ends with T4 polynucleotide kinase (Promega, http://www.promega.com), DNA was treated with λ-exonuclease to digest 5′ nicked DNA ends; replication intermediates protected by RNA primers are unaffected by this treatment. Primer extension reactions used ~500 ng of enriched replicating intermediates, 25 ng of radiolabeled primer (labeled using [γ^{32}P]-ATP, and T4 polynucleotide kinase) and 2 units of Deep Vent (exo-) DNA polymerase. After 30 cycles of reaction (60 s at 94 °C, 60 s at 70 °C, and 90 s at 72 °C), amplification products were separated on a 6% polyacrylamide gel under denaturing conditions. Radioactive material was detected using a Phosphorimager system (Amersham, http://www.gehealthcare.com). Control experiments using linearized plasmid DNA isolated from *E. coli* were performed under similar conditions.

Bioinformatics. Nucleotide representation disparities were calculated using either ORIGINX (http://www.cbs.dtu.dk/services/GenomeAtlas/suppl/origin) or ZPLOTTER (http://tubic.tju.edu.cn/zcurve) programs. Publicly available sequences released in June 2006 were used to calculate local minima and maxima in nucleotide disparities. DNA curvature was calculated using a BEND.IT server (http://hydra.icgeb.trieste.it/~kristian/dna/bend_it.html), which predicts in qualitative terms a curvature propensity of a given DNA sequence using DNase I based bendability parameters [41] and the consensus bendability scale [42]. The resulting data of each program were plotted using Origin Pro 7.5 software (OriginLab Corporation, http://www.originlab.com).

Supporting Information

Figure S1. *E. coli/H. volcanii* Shuttle Vectors with *ori-pHV1/4*

Shuttle vectors pTA351, pTA352, pTA354, and pTA356 are derived from pTA132, pTA133, pTA131, and pTA192, respectively [21], and feature the DNA replication origin from pHV1/4 isolated in the initial genetic screen.

Found at doi:10.1371/journal.pgen.0030077.sg001 (470 KB PDF).

Figure S2. Colony Size of *H. volcanii* Transformed with ARS Plasmids

H. volcanii H112 *(ΔpyrE2 ΔradA)* was transformed with pTA250 *(ori-pHV1/4),* pTA441 *(oriC1),* pTA612 *(oriC2),* pCN26 *(ori-pHV3),* or pCN27 *(ori-pHV4–2).* Cells were diluted 100-fold and 100 µl was plated on Hv-Ca, except for pCN27 where transformants were plated without dilution. Plates were incubated at 45 °C for 15 d before photographs were taken.

Found at doi:10.1371/journal.pgen.0030077.sg002 (3.3 MB PDF).

Table S1. Oligonucleotides

Found at doi:10.1371/journal.pgen.0030077.st001 (52 KB DOC).

Acknowledgments

We thank Moshe Mevarech for *H. volcanii* strains WR340 and WR480; Sam Haldenby for help with shuttle vector construction; Jonathan Badger for help in analysis of the *H. volcanii* genome; Stéphane Delmas for advice; Steve Bell for communicating results prior to publication; and Shiladitya DasSarma for constructive discussions.

Author contributions. CN, MH, HM, and TA conceived and performed the experiments. JAE sequenced the *H. volcanii* genome. CN, ALH, JAE, and HM analyzed the genome data. CN, MH, ALH, JAE, HM, and TA wrote the paper. CN and MH contributed equally to the work.

Funding. We are grateful to the BBSRC, MRC, and European Union Repbiotech project for funding. We are grateful to the Royal Society for a University Research Fellowship awarded to TA. HM also acknowledges the generous financial support from Fondation Bettencourt-Schueller and INSERM. The *H. volcanii* genome sequencing was supported by a National Science Foundation grant EF-024349 to JAE.

Competing interests. The authors have declared that no competing interests exist.

References

1. Stillman B (2005) Origin recognition and the chromosome cycle. FEBS Lett 579: 877–884.
2. Robinson NP, Bell SD (2005) Origins of DNA replication in the three domains of life. FEBS J 272: 3757–3766.
3. Barry ER, Bell SD (2006) DNA replication in the archaea. Microbiol Mol Biol Rev 70: 876–887.
4. Myllykallio H, Lopez P, Lopez-Garcia P, Heilig R, Saurin W, et al. (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. Science 288: 2212–2215.
5. Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R (2004) Three replication origins in *Sulfolobus* species: Synchronous initiation of chromosome replication and asynchronous termination. Proc Natl Acad Sci U S A 101: 7046–7051.
6. Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, et al. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell 116: 25–38.
7. Matsunaga F, Forterre P, Ishino Y, Myllykallio H (2001) In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proc Natl Acad Sci U S A 98: 11152–11157.
8. Capaldi SA, Berger JM (2004) Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. Nucleic Acids Res 32: 4821–4832.
9. Grainge I, Gaudier M, Schuwirth BS, Westcott SL, Sandall J, et al. (2006) Biochemical analysis of a DNA replication origin in the archaeon *Aeropyrum pernix*. J Mol Biol 363: 355–369.
10. Allers T, Mevarech M (2005) Archaeal genetics: The third way. Nat Rev Genet 6: 58–73.
11. Soppa J (2006) From genomes to function: Haloarchaea as model organisms. Microbiology 152: 585–590.
12. Baliga NS, Bonneau R, Facciotti MT, Pan M, Glusman G, et al. (2004) Genome sequence of *Haloarcula marismortui:* A halophilic archaeon from the Dead Sea. Genome Res 14: 2221–2234.
13. Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, et al. (2000) Genome sequence of *Halobacterium* species NRC-1. Proc Natl Acad Sci U S A 97: 12176–12181.
14. Charlebois RL, Schalkwyk LC, Hofman JD, Doolittle WF (1991) Detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. J Mol Biol 222: 509–524.
15. Berquist BR, DasSarma S (2003) An archaeal chromosomal autonomously replicating sequence element from an extreme halophile, *Halobacterium* sp. strain NRC-1. J Bacteriol 185: 5959–5966.
16. Grabowski B, Kelman Z (2003) Archaeal DNA replication: Eukaryal proteins in a bacterial context. Annu Rev Microbiol 57: 487–516.
17. Singleton MR, Morales R, Grainge I, Cook N, Isupov MN, et al. (2004) Conformational changes induced by nucleotide binding in Cdc6/ORC from *Aeropyrum pernix*. J Mol Biol 343: 547–557.
18. Bitan-Banin G, Ortenberg R, Mevarech M (2003) Development of a gene

knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. J Bacteriol 185: 772–778.
19. Charlebois RL, Lam WL, Cline SW, Doolittle WF (1987) Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaebacterium. Proc Natl Acad Sci U S A 84: 8530–8534.
20. Wendoloski D, Ferrer C, Dyall-Smith ML (2001) A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. Microbiology 147: 959–964.
21. Allers T, Ngo H, Mevarech M, Lloyd RG (2004) Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. Appl Environ Microbiol 70: 943–953.
22. Woods WG, Dyall-Smith ML (1997) Construction and analysis of a recombination-deficient *(radA)* mutant of *Haloferax volcanii*. Mol Microbiol 23: 791–797.
23. Cann IK, Komori K, Toh H, Kanai S, Ishino Y (1998) A heterodimeric DNA polymerase: Evidence that members of Euryarchaeota possess a distinct DNA polymerase. Proc Natl Acad Sci U S A 95: 14250–14255.
24. Komori K, Fujikane R, Shinagawa H, Ishino Y (2002) Novel endonuclease in Archaea cleaving DNA with various branched structure. Genes Genet Syst 77: 227–241.
25. Zhao A, Gray FC, Macneill SA (2006) ATP- and NAD-dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. Mol Microbiol 59: 743–752.
26. Guy CP, Bolt EL (2005) Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. Nucleic Acids Res 33: 3678–3690.
27. Lobry JR (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. Mol Biol Evol 13: 660–665.
28. Lopez P, Philippe H, Myllykallio H, Forterre P (1999) Identification of putative chromosomal origins of replication in Archaea. Mol Microbiol 32: 883–886.
29. Worning P, Jensen LJ, Hallin PF, Staerfeldt HH, Ussery DW (2006) Origin of replication in circular prokaryotic chromosomes. Environ Microbiol 8: 353–361.
30. Zhang R, Zhang CT (2005) Identification of replication origins in archaeal genomes based on the Z-curve method. Archaea 1: 335–346.
31. Kennedy SP, Ng WV, Salzberg SL, Hood L, DasSarma S (2001) Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. Genome Res 11: 1641–1650.
32. McLean MJ, Wolfe KH, Devine KM (1998) Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J Mol Evol 47: 691–696.
33. Gerbi SA, Bielinsky AK (1997) Replication initiation point mapping. Methods 13: 271–280.
34. Zhang R, Zhang CT (2003) Multiple replication origins of the archaeon *Halobacterium* species NRC-1. Biochem Biophys Res Commun 302: 728–734.
35. Donato JJ, Chung SC, Tye BK (2006) Genome-wide hierarchy of replication origin usage in *Saccharomyces cerevisiae*. PLoS Genet 2: e141. doi:10.1371/journal.pgen.0020141
36. Kolstø AB (1999) Time for a fresh look at the bacterial chromosome. Trends Microbiol 7: 223–226.
37. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, et al. (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature 406: 477–483.
38. Sambrook J, Russell DW (2001) Molecular cloning: A laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. 999 p.
39. Guy CP, Haldenby S, Brindley A, Walsh DA, Briggs GS, et al. (2006) Interactions of RadB, a DNA repair protein in Archaea, with DNA and ATP. J Mol Biol 358: 46–56.
40. Mullakhanbhai MF, Larsen H (1975) *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. Arch Microbiol 104: 207–214.
41. Brukner I, Sanchez R, Suck D, Pongor S (1995) Sequence-dependent bending propensity of DNA as revealed by DNase I: Parameters for trinucleotides. EMBO J 14: 1812–1818.
42. Gabrielian A, Pongor S (1996) Correlation of intrinsic DNA curvature with DNA property periodicity. FEBS Lett 393: 65–68.